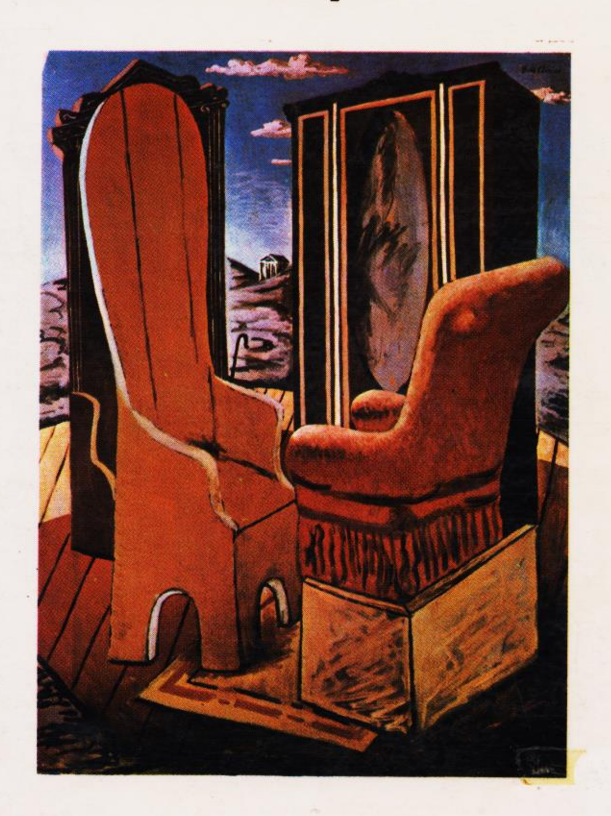
LA FILOSOFIA Y EL ESPEJO DE LA NATURALEZA

Richard Rorty



CATEDRA

La filosofía y el espejo de la naturaleza

Colección Teorema

Serie mayor

Richard Rorty

La filosofía y el espejo de la naturaleza

SEGUNDA EDICIÓN

CATEDRA
TEOREMA

Título original de la obra:
Philosophy and the Mirror of Nature

Traducción de Jesús Fernández Zulaica

Diseño de cubierta: Diego Lara

Cubierta: Manuel Luca de Tena

Ediciones Cátedra, S. A., 1989
Josefa Valcárcel, 27. 28027-Madrid
Depósito legal: M. 41.977-1989
ISBN: 84-376-0432-X
Printed in Spain
Impreso en Lavel
Los Llanos, nave 6. Humanes (Madrid)

Índice

TERCERA PARTE
La Filosofía

To
M. V. R.

Cuando pensamos en el futuro del mundo, siempre tenemos la idea de que está en el lugar donde debería estar si siguiera moviéndose tal como lo vemos moverse ahora. No nos damos cuenta de que no se mueve en línea recta, sino curva, y que su dirección cambia constantemente.

¿No ha progresado nada la filosofía? Si alguien rasca donde pica, ¿debe considerarse eso como progreso? Si *no,* ¿significa esto que no fue un acto de rascar auténtico? ¿Ni un picor auténtico? ¿No podría prolongarse durante mucho tiempo esta respuesta al estímulo, antes de que se encontrara un remedio para el picor?

Wenn wir an die Zukunft der Welt denken, so meinen wir immer den Ort, wo sie sein wird, wenn sie so weiter läuft, wie wir sie jetzt laufen sehen, und denken nicht, dass sie nicht gerade läuft, sondern in einer Kurve, und ihre Richtung sich konstant ändert. (Ludwig Wittgenstein, *Vermischte Bemerkungen,* Frankfurt, 1977, página 14.)

Die Philosophie hat keinen Fortschritt gemacht? Wenn Einer kratzt, wo es ihn juckt, muss ein Fortschritt zu sehen sein? Ist es sonst kein echtes Kratzen, oder kein echtes Jucken? Und kann nicht diese Reaktion auf die Reizung lange Zeit so weitergehen, ehe ein Mittel gegen das Jucken gefunden wird? *(ibíd.,* páginas 163-164).

PREFACIO

Casi desde el mismo momento en que comencé a estudiar filosofía me impresionó la forma en que los problemas filosóficos aparecían o cambiaban de forma, como consecuencia de la adopción de nuevas suposiciones o vocabularios. De Richard Mckeon y Robert Brumbaugh aprendí a considerar la historia de la filosofía como una serie, no de soluciones alternativas a los mismos problemas sino de conjuntos de problemas muy diferentes. De Rudolph Carnap y Carl Hempel aprendí cómo es posible que los pseudo-problemas aparezcan como tales cuando se vuelven a formular en el modo formal del habla. De Charles Hartshorne y Paul Weiss aprendí cómo se podía manifestar su carácter traduciéndolos a la terminología de Whitehead o Hegel. Tuve la gran suerte de contar con estos hombres entre mis maestros, pero, para bien o para mal, consideré que todos ellos decían lo mismo: que un «problema filosófico» era producto de la adopción inconsciente de suposiciones incorporadas al vocabulario en que se formulaba el problema —suposiciones que había que cuestionar antes de abordar seriamente el mismo problema.

Algo más tarde, comencé a leer las obras de Wilfrid Sellars. Me pareció que el ataque de Sellars al «Mito de lo Dado» daba carácter dudoso a las suposiciones existentes en el trasfondo de la mayor parte de la filosofía moderna. Y algo más adelante, comencé a tomarme en serio el planteamiento escéptico de Quine ante la distinción lenguaje-hecho, y a tratar de combinar el punto de vista de Quine con el de Sellars. Desde entonces, estoy tratando de aislar el mayor número posible de suposiciones existentes tras la problemática de la filosofía moderna, con la esperanza de generalizar y ampliar las críticas de Sellars y Quine al empirismo tradicional. Volviendo a estas suposiciones, y dejando claro que son opciones, pensaba yo, se conseguirían resultados «terapéuticos», de forma parecida a como ocurrió con la disolución original de los problemas de los manuales clá-

sicos, promovida por Carnap. El presente libro es fruto de ese intento.

El proceso de elaboración del mismo ha sido largo. La Universidad de Princeton es especialmente generosa en relación con el tiempo dedicado a la investigación y los años sabáticos, por lo que me resulta violento confesar que sin la ayuda adicional del «American Council of Learned Societies» (Consejo americano de sociedades culturales) y de la Fundación en memoria de John Simon Guggenheim, probablemente no habría llegado a escribirlo nunca. Comencé a esbozar sus líneas generales cuando estaba disfrutando de una beca del ACLS en 1969-1970, y escribí la mayor parte de la primera redacción al recibir una beca Guggenheim en 1973-1974. Estoy sumamente agradecido a las tres instituciones por la ayuda prestada.

Muchas personas —alumnos de Princeton y otras universidades, asistentes a diversas conferencias, colegas y amigos— han leído o escuchado diversas versiones de varias secciones del libro. He introducido numerosos cambios tanto de contenido como de estilo en respuesta a sus objeciones y por ello les estoy muy agradecido. Siento que mi memoria sea tan mala que no puedo enumerar ni siquiera los casos más importantes de dicha ayuda, pero espero que los lectores que me han ayudado reconozcan a lo largo de la obra la influencia positiva de sus comentarios. No obstante, quiero manifestar mi agradecimiento a dos personas —Michael Williams y Richard Bernstein— que han realizado comentarios muy valiosos sobre la penúltima versión de la obra completa igual que un lector anónimo de la Princeton University Press. Estoy muy agradecido a Raymond Geuss, David Hoy y Jeffrey Stout, que me han dedicado parte de su tiempo y me han ayudado a resolver las dudas de última hora sobre el capítulo final.

Para terminar, me gustaría expresar mi agradecimiento a Laura Bell, Pearl Cavanaugh, Lee Ritins, Carol Roan, Sanford Thatcher, Jean Toll y David Velleman por haberme ayudado pacientemente a transformar el borrador que yo había escrito en un volumen impreso.

* * *

Algunos fragmentos del Capítulo IV habían aparecido en *Neue Hefte für Philosophie* 14 (1978). Partes del Capítulo V, en *Body, Mind and Method: Essays in Honor of Virgil C. Aldrich,* dirigido por Donald F. Gustafson y Bangs L. Tapscott (Dordrecht, 1979). Otras partes del mismo capítulo habían sido publicadas en *Philosophical Studies,* 31 (1977). Partes del Capítulo VII, en *Acta philosophica Fennica,* 1979. Quiero dar las gracias a los directores y editores que han autorizado la reimpresión de dicho material.

INTRODUCCIÓN

Por lo general, los filósofos piensan que su disciplina se ocupa de problemas perennes, eternos —problemas que surgen en el momento en que se reflexiona. Algunos de ellos se refieren a la diferencia entre los seres humanos y los demás seres, y se cristalizan en preguntas sobre la relación entre la mente y el cuerpo. Otros problemas hacen referencia a la legitimación de las ansias de conocer, y se cristalizan en preguntas sobre los «fundamentos» del conocimiento. Descubrir estos fundamentos es descubrir algo sobre la mente, y al revés. Por eso, la filosofía en cuanto disciplina se considera a sí misma como un intento de confirmar o desacreditar las pretensiones de conocimiento que se dan en la ciencia, en la moralidad, en el arte o en la religión. Trata de hacerlo basándose en su comprensión especial de la naturaleza del conocimiento y de la mente. La filosofía puede tener carácter de fundamento en relación con el resto de la cultura, pues ésta es la acumulación de las pretensiones de conocimiento, y la filosofía debe juzgarlas. Puede hacerlo porque comprende los fundamentos del conocimiento, y encuentra estos fundamentos en un estudio del hombre-en-cuanto-ser-que-conoce, de los «procesos mentales» o de la «actividad de representación» que hacen posible el conocimiento. Saber es representar con precisión lo que hay fuera de la mente; entender de esta manera la posibilidad y naturaleza del conocimiento es entender la forma en que la mente es capaz de reconstruir tales representaciones. La preocupación fundamental de la filosofía es ser una teoría general de la representación, una teoría que divida la cultura en áreas que representen bien la realidad, otras que la representen menos bien y otras que no la representen en absoluto (a pesar de su pretensión de hacerlo).

La idea de una «teoría del conocimiento» basada en una comprensión de los «procesos mentales» es producto del siglo XVII, y sobre todo de Locke. La idea de «la mente» en cuanto entidad en

la que ocurren los «procesos» aparece en ese mismo periodo especialmente en las obras de Descartes. Al siglo XVIII, y a Kant de una forma especial, debemos la idea de la filosofía en cuanto tribunal de la razón pura, que confirma o rechaza las pretensiones del resto de la cultura, pero esta idea kantiana presuponía un asentimiento general a las ideas de Locke sobre los procesos mentales y a las de Descartes sobre la sustancia mental. En el siglo XIX, la idea de la filosofía en cuanto disciplina fundamental que «sirve de base» a las pretensiones de conocimiento se vio consolidada en los autores neo-kantianos. Las protestas formuladas ocasionalmente contra esta concepción de la cultura como si tuviera necesidad de una «base» y contra las pretensiones de que esta tarea fuera realizada por una teoría del conocimiento (por ejemplo, en Nietzsche y William James) pasaron prácticamente desapercibidas. Para los intelectuales, la «filosofía» se convirtió en un sustituto de la religión. Era el área de la cultura donde se tocaba fondo, donde se encontraban el vocabulario y las convicciones que permitían explicar y justificar las actividades propias *en cuanto* intelectual, y descubrir, por tanto, el significado de la propia vida.

A comienzos de nuestro siglo, esta pretensión recibió la confirmación de los filósofos (especialmente Russell y Husserl) que tenían interés en que la filosofía siguiera siendo «rigurosa» y «científica». Pero en su voz se apreciaba cierto sentimiento de desesperación, pues para esas fechas el triunfo de lo secular sobre las reivindicaciones de la religión era casi total. De esta manera, el filósofo no podía considerarse por más tiempo en la vanguardia intelectual, o como protector de los hombres ante las fuerzas de la superstición. Además, en el curso del siglo XIX, había surgido una nueva forma de cultura —la cultura del hombre de letras, el intelectual que escribía poemas y novelas y tratados políticos, y críticas a los poemas y novelas y tratados de otros autores. Descartes, Locke y Kant habían escrito en un periodo en el que la secularización de la cultura comenzaba a ser una posibilidad gracias al triunfo de la ciencia natural. Pero, para comienzos del siglo XX, los científicos habían llegado a separarse de la mayoría de los intelectuales tanto como los teólogos. Poetas y novelistas habían ocupado el lugar de los predicadores y filósofos en cuanto maestros morales de la juventud. El resultado fue que cuanto más «científica» y «rigurosa» se hacía la filosofía, menos tenía que ver con el resto de la cultura y más absurdas parecían sus pretensiones tradicionales. Los intentos de los filósofos analíticos y de los fenomenólogos por «fundamentar» unas cosas y «criticar» otras no encontraron eco en aquellos cuyas acti-

vidades trataban de fundamentar o criticar. Los que necesitaban una ideología o una auto-imagen hicieron caso omiso de la filosofía en general.

Este es el telón de fondo sobre el que hay que proyectar la obra de los tres filósofos más importantes de nuestro siglo —Wittgenstein Heidegger y Dewey. Todos ellos trataron, en un primer momento, de encontrar un nuevo modo de dar a la filosofía carácter «básico» —una nueva manera de formular un contexto definitivo del pensamiento. Wittgenstein intentó construir una nueva teoría de la representación que no tuviera nada que ver con el mentalismo; Heidegger, un nuevo conjunto de categorías filosóficas que no tuviera nada que ver con la ciencia, la epistemología o la búsqueda cartesiana de la certeza; y Dewey, una versión naturalizada de la visión hegeliana de la historia. Cada uno de ellos terminó considerando que sus primeros esfuerzos habían estado mal dirigidos, que eran un intento de mantener una cierta concepción de la filosofía después de haber rechazado las nociones necesarias para dar contenido a tal concepción (las ideas del siglo XVIII sobre el conocimiento y la mente). Todos ellos, en sus obras posteriores, se emanciparon de la concepción kantiana de la filosofía en cuanto disciplina básica, y se dedicaron a ponernos en guardia frente a las mismas tentaciones en que ellos habían caído. Por eso, sus escritos últimos son más terapéuticos que constructivos, más edificantes que sistemáticos, orientados a hacer que el lector se cuestione sus propios motivos para filosofar más que a presentarle un nuevo programa filosófico.

Wittgenstein, Heidegger y Dewey están de acuerdo en que hay que abandonar la noción del conocimiento en cuanto representación exacta, que resulta posible gracias a procesos mentales especiales e inteligible gracias a una teoría general de la representación. Los tres consideran que se deben descartar las nociones de «fundamentos del conocimiento» y de la filosofía en cuanto centrada en el intento cartesiano de dar respuesta al escéptico epistemológico. Además, prescinden de la idea de «la mente» común a Descartes, Locke y Kant —en cuanto tema especial de estudio, situada en el espacio interior y dotada de elementos o procesos que posibilitan el conocimiento. Esto no quiere decir que tengan *otras* «teorías del conocimiento» o «filosofías de la mente». Rechazan la epistemología y la metafísica en cuanto disciplinas posibles. Hablo de «dejar de lado» y «rechazar» en vez de «presentar argumentos en contra» porque su actitud hacia la problemática tradicional es como la de los filósofos del siglo XVII hacia la problemática escolástica. No suelen entretenerse en descubrir proposiciones erróneas o falsos argumentos en las obras

de sus predecesores (aunque lo hagan de vez en cuando). Lo que hacen, más bien, es vislumbrar la posibilidad de una forma de vida intelectual en la que el vocabulario de la reflexión filosófica heredado del siglo XVII parecería tan fuera de lugar como se lo había parecido a la Ilustración el vocabulario filosófico del siglo XIII. Afirmar la posibilidad de una cultura post-kantiana, en la que no haya ninguna disciplina global que legitime o sirva de base a las otras, no significa necesariamente atacar ninguna doctrina kantiana concreta, lo mismo que vislumbrar la posibilidad de una cultura en que la religión no existiera o no tuviera ninguna conexión con la ciencia o la política, no implicaba necesariamente atacar la afirmación de Tomás de Aquino de que es posible demostrar la existencia de Dios con la razón natural. Wittgenstein, Heidegger y Dewey nos han introducido en un periodo de filosofía «revolucionaria» (en el sentido de la ciencia «revolucionaria» de Kuhn). Introduciendo nuevos mapas del terreno (a saber, de todo el panorama de las actividades humanas) en que no aparecen los rasgos que anteriormente parecían tener carácter dominante.

Este libro es una visión de conjunto de algunos avances recientes dentro de la filosofía, especialmente de la filosofía analítica, desde el punto de vista de la revolución anti-cartesiana y anti-kantiana que acabo de describir. El objetivo de la obra es acabar con la confianza que el lector pueda tener en «la mente» en cuanto algo sobre lo que se debe tener una visión «filosófica», en el «conocimiento» en cuanto algo que debe ser objeto de una «teoría» y que tiene «fundamentos», y en la «filosofía» tal como se viene entendiendo desde Kant. El lector que vaya buscando una nueva teoría sobre cualquiera de los temas mencionados se sentirá defraudado. Aunque trato de las «soluciones al problema mente-cuerpo» no lo hago con la intención de proponer una, sino de ilustrar por qué pienso que no existe ese problema. Igualmente, aunque me ocupo de las «teorías de la referencia» no presento ninguna, sino que me limito a ofrecer sugerencias sobre las razones por las que la búsqueda de dicha teoría está mal enfocada. El libro, como la mayoría de las obras de los autores que más admiro, es terapéutico, más que constructivo. La terapia ofrecida depende, no obstante, de los esfuerzos constructivos de los mismos filósofos analíticos cuyo marco de referencia estoy intentando cuestionar. Así, la mayoría de las críticas concretas de la tradición están tomadas de filósofos tan sistemáticos como Sellars, Quine, Davidson, Ryle, Malcolm, Kuhn y Putnam.

Es tanto lo que debo a estos filósofos en relación con los medios que utilizo como lo que debo a Wittgenstein, Heidegger y

Dewey en relación con los fines a que se aplican. Espero convencer al lector de que la dialéctica existente dentro de la filosofía analítica, que ha llevado a la filosofía de la mente desde Broad hasta Smart, a la filosofía del lenguaje desde Frege a Davidson, a la epistemología desde Russell hasta Sellars, y a la filosofía de la ciencia desde Carnap hasta Kuhn, debe hacerse progresar algunos pasos más. Estos pasos adicionales, pienso yo, nos colocarán en situación de criticar la misma idea de «filosofía analítica», y hasta de la misma «filosofía» tal como ha sido entendida desde la época de Kant.

Desde el punto de vista que estoy adoptando, la diferencia entre la filosofía «analítica» y las otras clases de filosofía carece relativamente de importancia —sería en realidad una cuestión de estilo y tradición más que de una diferencia de «método» o de primeros principios. La razón por la que el presente libro está escrito fundamentalmente con el vocabulario de los filósofos analíticos contemporáneos y hace referencia a los problemas examinados en las obras de carácter analítico, es exclusivamente autobiográfica. Son el vocabulario y las obras con que estoy más familiarizado y a los que debo la comprensión que yo pueda tener de los problemas filosóficos. Si estuviera igualmente familiarizado con otras formas actuales de escribir filosofía, la obra resultaría más completa y útil, aunque todavía más larga de lo que es. Tal como yo entiendo las cosas, el tipo de filosofía que procede de Russell y Frege, como la fenomenología husserliana clásica, es sencillamente un intento más de colocar a la filosofía en la situación que Kant deseaba para ella —la de juzgar las demás áreas de la cultura basándose en su especial conocimiento de los «fundamentos» de estas áreas. La filosofía «analítica» es una nueva variante de la filosofía kantiana, una variante que se caracteriza principalmente por considerar que la representación es lingüística más que mental, y que la filosofía del lenguaje, más que «crítica transcendental» o psicología, es la disciplina que presenta los «fundamentos del conocimiento». Esta insistencia en el lenguaje, como insistiré en los capítulos cuatro y seis, no cambia esencialmente la problemática cartesiano-kantiana, y por tanto no da a la filosofía una nueva auto-imagen. La filosofía analítica sigue empeñada en la construcción de un marco de referencia permanente y neutral para la investigación, y por tanto para toda la cultura.

Lo que vincula la filosofía contemporánea con la tradición de Descartes-Locke-Kant es la idea de que la actividad humana (y la investigación, la búsqueda del conocimiento, en especial) se produce dentro de un marco que se puede aislar antes de la conclusión de la

investigación —un conjunto de presuposiciones que se pueden descubrir a priori. La idea de que existe dicho marco sólo tiene sentido si pensamos que viene impuesto por la naturaleza del sujeto que conoce, por la naturaleza de sus facultades o por la naturaleza del medio dentro del cual actúa. La misma idea de la «filosofía» como algo distinto de la «ciencia» no tendría mucho sentido sin la afirmación cartesiana de que mirando hacia dentro podemos dar con la verdad ineluctable, y la afirmación kantiana de que esta verdad impone límites a los posibles resultados de la investigación empírica. La idea de que pueda haber eso que se llama «fundamentos del conocimiento» (de *todo* conocimiento —en todos los campos, del pasado, presente y futuro) o una «teoría de la representación» (de *toda* representación, en los vocabularios conocidos y en los que ahora no podemos ni imaginar) depende de la suposición de que existe semejante constricción a priori. Si tenemos una concepción deweyana del conocimiento, como lo que creemos justificadamente, no nos imaginaremos que existan limitaciones duraderas a lo que puede figurar como conocimiento, pues veremos la «justificación» como un fenómeno social más que como una transacción entre «el sujeto que conoce» y «la realidad». Si tenemos una concepción wittgensteiniana del lenguaje, en cuanto instrumento más que como espejo, no buscaremos las condiciones necesarias de posibilidad de la representación lingüística. Si tenemos una concepción heideggeriana de la filosofía, consideraremos que el intento de convertir la naturaleza del sujeto que conoce en fuente de verdades necesarias es uno de los intentos descaminados de recurrir a una cuestión «técnica» y determinada en vez de a la apertura a lo desconocido que fue la tentación inicial que nos indujo a pensar.

Una forma de ver cómo se encaja la filosofía analítica dentro del esquema cartesiano-kantiano tradicional es ver la filosofía tradicional como un intento de escapar a la historia —un intento de encontrar condiciones ahistóricas de cualquier posible cambio histórico. Desde esta perspectiva, el mensaje común de Wittgenstein, Dewey y Heidegger es historicista. Todos ellos nos recuerdan que las investigaciones de los fundamentos del conocimiento o de la moralidad o del lenguaje o de la sociedad quizá no sean más que una apologética, un intento de eternizar un determinado juego lingüístico, práctica social o auto-imagen contemporáneos. También la conclusión de este libro es historicista, y las tres partes en que se divide tratan de colocar en su perspectiva histórica las nociones de «mente», de «conocimiento» y de «filosofía», respectivamente. La Parte I se ocupa de la filosofía de la mente, y en el capítulo primero trato de

hacer ver que las llamadas intuiciones que están tras el dualismo cartesiano tienen origen histórico. En el capítulo segundo, intento mostrar cómo cambiarían estas intuiciones si los métodos psicológicos dieran paso a métodos fisiológicos de predicción y control.

La Parte II trata de la epistemología y de los intentos recientes de encontrar «nuevos temas» de la epistemología. El capítulo tercero describe la génesis de la noción de «epistemología» en el siglo XVII, y su conexión con las nociones cartesianas de la «mente» expuestas en el capítulo primero. Presenta la «teoría del conocimiento» como una noción basada en la confusión existente entre la justificación de las pretensiones de conocimiento y su explicación causal —es decir, en términos aproximados, entre las prácticas sociales y los procesos psicológicos que se postulan. El capítulo cuarto es el más importante del libro: en él se presentan las ideas que han sido la causa de que haya llegado a escribirse. Son las ideas de Sellars y de Quine, y en este capítulo interpreto el ataque de Sellars a «lo dado» y el de Quine a la «necesidad» como los pasos fundamentales que han minado la posibilidad de una «teoría del conocimiento». El holismo y pragmatismo que comparten ambos filósofos, y que comparten a su vez con el Wittgenstein de los últimos años, son las líneas de pensamiento dentro de la filosofía analítica que deseo ampliar. Señalo que, cuando se amplían de una determinada manera, nos permiten ver la verdad no como «la representación exacta de la realidad» sino como «lo que nos es más conveniente creer», utilizando la expresión de James. O, dicho menos provocativamente, nos demuestran que la idea de «representación exacta» no pasa de ser un cumplido automático y sin contenido que hacemos a las creencias que consiguen ayudarnos a hacer lo que queremos hacer. En los capítulos cinco y seis examino y critico lo que considero como intentos reaccionarios de tratar la psicología empírica o la filosofía del lenguaje como los «nuevos temas» de la epistemología. Mantengo que sólo la idea del conocimiento como «exactitud de la representación» nos convence de que el estudio de los procesos psicológicos o del lenguaje —en cuanto medios de representación— puede conseguir lo que no llegó a hacer la epistemología. La conclusión de la parte II en conjunto es que la noción del conocimiento en cuanto acumulación de representaciones precisas es opcional —que puede reemplazarse por una concepción pragmática del conocimiento que elimine el contraste griego entre contemplación y acción, entre representar el mundo y enfrentarse con él. Una época histórica dominada por las metáforas oculares griegas puede dar paso a otra en que el vocabulario filosófico que incorpore

estas metáforas parezca tan extraño como el vocabulario animista de los tiempos pre-clásicos.

En la parte III abordo más explícitamente la idea de «filosofía». El capítulo siete interpreta la distinción tradicional entre la búsqueda del «conocimiento objetivo» y otras áreas, menos privilegiadas, de la actividad humana como una mera distinción entre «discurso normal» y «discurso anormal». El discurso normal (generalización de la idea de «ciencia normal» de Kuhn) es todo discurso (científico, político, teológico, etc.) que encarne los criterios aceptados para llegar a un acuerdo; discurso anormal es el que no contenga estos criterios. Mi punto de vista es que el intento (que ha sido característico de la filosofía tradicional) de explicar la «racionalidad» y la «objetividad» en términos de condiciones de representación precisa es un esfuerzo engañoso de eternizar el discurso normal del momento, y que, desde los griegos, la auto-imagen de la filosofía ha estado dominada por este intento. En el capítulo ocho recurro a algunas ideas tomadas de Gadamer y Sartre para desarrollar el contraste que existe entre filosofía «sistemática» y «edificante», y mostrar cómo se relaciona con la filosofía «normal» la filosofía «anormal» que no se adapta a la matriz tradicional cartesiano-kantiana. Presento a Wittgenstein, Heideger y Dewey como filósofos cuyo objetivo es edificar —ayudar a sus lectores, o a la sociedad en su conjunto, a liberarse de actitudes y vocabularios caducos, y no ofrecerles una «base» para las intuiciones y costumbres del presente.

Espero que lo que vengo diciendo haya aclarado por qué he elegido el título «La filosofía y el espejo de la naturaleza». Son imágenes más que proposiciones, y metáforas más que afirmaciones, lo que determina la mayor parte de nuestras convicciones filosóficas. La imagen que mantiene cautiva a la filosofía tradicional es la de la mente como un gran espejo, que contiene representaciones diversas —algunas exactas, y otras no— y se puede estudiar con métodos puros, no empíricos. Sin la idea de la mente como espejo, no se habría abierto paso la noción del conocimiento como representación exacta. Sin esta última idea, no habría tenido sentido la estrategia común de Descartes y Kant —obtener representaciones más exactas inspeccionando, reparando y limpiando el espejo, por así decirlo. Sin esta estrategia, habrían carecido de sentido las afirmaciones más recientes de que la filosofía podría consistir en el «análisis conceptual» o en el «análisis fenomenológico» o en la «explicación de los significados» o en el examen de «la lógica de nuestro lenguaje» o de «la estructura de la actividad constituyente de la conciencia». Este tipo de afirmaciones son las que ridiculizaba Wittgenstein en

sus *Investigaciones filosóficas,* y, siguiendo las orientaciones de Wittgenstein, la filosofía analítica ha progresado hacia la postura «postpositivista» que adopta en la actualidad. Pero el talento de Wittgenstein para desmontar las imágenes cautivadoras debe complementarse con la conciencia histórica —la conciencia de la fuente de todas las imágenes especulares— y ésa es, en mi opinión, la mayor aportación de Heidegger. La forma en que Heidegger vuelve a contar la historia de la filosofía nos permite ver los comienzos de las imágenes cartesianas en los griegos así como las metamorfosis de estas imágenes durante los tres últimos siglos. De esta forma nos permite «distanciarnos» de la tradición. Sin embargo, ni Heidegger ni Wittgenstein nos hacen ver el fenómeno histórico de las imágenes basadas en el espejo, la historia de la dominación de la mente de Occidente por las metáforas oculares, dentro de una perspectiva social. Los dos se centran en el individuo singularmente favorecido más que en la sociedad —en las posibilidades de mantenerse al margen de la banal auto-decepción característica de las últimas fechas de una tradición decadente. Por el contrario, Dewey, aunque no tenía ni la agudeza dialéctica de Wittgenstein ni la erudición histórica de Heidegger, escribió sus ataques a las tradicionales imágenes del espejo partiendo de la visión de un nuevo tipo de sociedad. En su sociedad ideal, la cultura ya no está dominada por el ideal del conocimiento objetivo sino del desarrollo estético. En dicha cultura, como él decía, las artes y las ciencias serían «las flores naturales de la vida». Tengo la esperanza de que estemos ahora en condiciones de ver las acusaciones de «relativismo» e «irracionalismo» presentadas en su momento contra Dewey como simples reflejos defensivos e inconscientes de la tradición filosófica que él atacaba. Dichas acusaciones no tienen ningún peso si se toman en serio las críticas que él, Wittgenstein y Heidegger hacen a las imágenes basadas en la comparación con el espejo. Este libro no tiene mucho que añadir a dichas críticas, pero espero que presente algunas de ellas de forma que contribuya a romper la coraza de la convención filosófica que Dewey esperaba, en vano, reducir a añicos.

PRIMERA PARTE:

Nuestra esencia de vidrio

Capítulo primero

La invención de la mente

1. Criterios de lo mental

Los estudios de la filosofía de la mente suelen comenzar suponiendo que todos hemos sabido siempre dividir el mundo en dos categorías: lo mental y lo físico, y que esta distinción es de sentido común e intuitiva aun cuando la existente entre lo material y lo inmaterial sea filosófica y desconcertante. Así, cuando Ryle sugiere que hablar de entidades mentales es hablar de disposiciones al comportamiento, o cuando Smart sugiere que es hablar de estados neurales, tienen dos cosas en contra. Si el conductismo o el materialismo están en lo cierto, ¿por qué debe haber esa distinción intuitiva?

Parece que no tenemos ninguna duda de que los dolores, talantes, imágenes y oraciones que «destellan ante la mente», los sueños, alucinaciones, creencias, actitudes, deseos e intenciones, todos ellos figuran como «mentales», mientras que las contracciones del estómago que causan el dolor, los procesos nerviosos que lo acompañan, y todo lo demás a lo que se puede atribuir una localización clara dentro del cuerpo figuran como no-mentales. Nuestra clasificación no se plantea ninguna duda y sugiere no sólo que tenemos una intuición clara de lo que es la «mentalidad», sino que tiene algo que ver con la no-espacialidad y con la noción de que aunque se destruyera el cuerpo, las entidades o estados mentales podrían seguir perdurando de alguna manera. Aun cuando rechacemos la noción de «sustancia mental», aun cuando prescindamos de la noción de *res cogitans* en cuanto sujeto de predicación, parecemos capaces de distinguir entre la mente y el cuerpo, y hacerlo en una forma más o menos cartesiana.

Estas supuestas intuiciones sirven para mantener con vida algo parecido al dualismo cartesiano. Los filósofos post-wittgensteinianos que se oponen al conductismo y al materialismo suelen reconocer a Wittgenstein y Strawson que, en cierto sentido, no hay nada más que el organismo humano, y que debemos rechazar la noción de que este organismo esté compuesto de un poco de *res cogitans* asociada no-espacialmente con un poco de *res extensa.* Pero, dicen, sigue en pie la intuición cartesiana de que la distinción entre lo mental y lo físico es insuperable con medios empíricos, de que un estado mental no se parece más a una disposición que a una neurona, y de que ningún descubrimiento científico puede revelar una identidad. Esta intuición les parece lo suficientemente fuerte como para abrir un abismo infranqueable. Pero estos filósofos neo-dualistas se ven comprometidos por sus propias conclusiones, pues aunque sus intuiciones metafísicas puedan parecer cartesianas, no tienen claro si les está permitido *tener* eso que se llama «intuiciones metafísicas». Suelen mostrarse descontentos ante la idea de un método de tener conocimiento del mundo que sea anterior a la ciencia empírica y ésta no pueda alcanzar.

En esta situación, el dualista siente la tentación de pasarse a la lingüística y comenzar a hablar de «vocabularios diferentes» o «descripciones alternativas». Esta jerga quiere dar a entender que la intuición dualista en cuestión no es más que una de las diferencias entre las formas de hablar sobre el mismo fenómeno, y por tanto parece conducirnos desde algo como el dualismo a algo como la teoría del doble aspecto de Spinoza. Pero la pregunta «dos descripciones ¿de *qué?*» hace que sea difícil mantenerse en esa posición. La respuesta «dos descripciones de organismos» parece correcta mientras no preguntemos: «¿Son físicos los organismos?» o «¿Hay en los organismos, incluso en los organismos humanos, algo más que las disposiciones actuales y posibles de sus partes?» Los neo-dualistas suelen conceder con facilidad a Ryle un gran número de estados mentales, y dicen que las creencias, deseos, actitudes e intenciones (por no hablar de las destrezas, virtudes y humores) no pasan de ser meras formas de hablar sobre los organismos, sus partes y los movimientos actuales y posibles de dichas partes. (Pero pueden insistir, siguiendo a Brentano y a Chisholm, en que no se pueden dar las condiciones necesarias y suficientes de Ryle). Ahora bien, cuando hablan de los dolores, imágenes mentales y pensamientos producidos —estados mentales de breve duración que, por así decirlo, se parecen más a hechos que a disposiciones— comienzan a dudar. Y tienen razones para ello. La diferencia entre dualismo y materialis-

mo desaparecería si en algún momento dijeran que describir un organismo en cuanto sometido al dolor no es más que una forma de hablar sobre el estado de sus partes. Estas partes, téngase en cuenta, han de ser *físicas,* pues una vez que hemos «kantizado» y «strawsonizado» a Descartes ya no tendrá sentido la noción de «parte mental». ¿Qué más podría pedir un defensor de la identidad mente-cuerpo que la admisión de que hablar de cómo se *siente* alguien es solamente una forma alternativa de informar de cómo *están* las porciones correspondientes de su anatomía (probablemente las neuronas)?

Tenemos así el siguiente dilema: o los neo-dualistas deben elaborar una explicación epistemológica de cómo sabemos a priori que las cosas se dividen en dos especies ontológicas irreductiblemente distintas, o de lo contrario deben encontrar alguna forma de expresar su dualismo que no se base ni en la noción de «abismo ontológico» ni en la de «descripción alternativa». Pero antes de buscar formas de resolver este dilema, deberíamos considerar más atentamente la noción de «especie ontológica» o «abismo ontológico». ¿Qué clase de noción es ésta? ¿Tenemos otros ejemplos de abismos ontológicos? ¿Hay algún otro caso en que sepamos a priori que ninguna investigación empírica puede identificar las dos entidades? Sabemos, quizá, que ninguna investigación empírica puede identificar dos entidades espacio-temporales que tengan localizaciones diferentes, pero ese conocimiento parece demasiado trivial para ser relevante. ¿Existe algún otro caso en el que sepamos a priori algo sobre las clases ontológicas naturales? Los únicos ejemplos que se me ocurren son los de las distinciones entre finito e infinito, entre humano y divino, y entre particular y universal. Nada, intuimos, podría atravesar *esas* líneas divisorias. Pero estos ejemplos no resultan demasiado útiles. Nos sentimos inclinados a decir que no sabemos qué significaría que existiera algo infinito. Si tratamos de clarificar la noción ortodoxa de «lo divino» parece que tenemos o una concepción meramente negativa o una que se explicaría haciendo referencia a las nociones de «infinitud» e «inmaterialidad». Como la referencia a la infinitud explica lo oscuro recurriendo a lo más oscuro, lo único que nos queda es la inmaterialidad. Tenemos la vaga confianza de que si *pudiera* existir lo infinito, sólo podría estar ejemplificado, como lo universal, por lo inmaterial. Si tiene algún sentido hablar de la existencia de los universales, parecería que deben existir inmaterialmente y *ésa* es la razón por la que no se pueden identificar nunca con los particulares espacio-temporales. Pero, ¿qué significa «inmaterial»? ¿Es lo mismo que «mental»?

Aun cuando resulte difícil ver en la noción de ser «físico» algo más que ser «material» o «espacio-temporal», no está claro que «mental» e «inmaterial» sean sinónimos. Si lo fueran, las disputas sobre la categoría de los universales producidas entre conceptualistas y realistas parecerían todavía más absurdas. Sin embargo, lo contrario de «mental» es «físico» y lo contrario de «inmaterial» es «material». «Físico» y «material» parecen sinónimos. ¿Cómo es posible que dos conceptos distintos tengan opuestos sinónimos?

Podemos sentir aquí la tentación de recurrir a Kant y explicar que lo mental es temporal pero no espacial, mientras que lo inmaterial —el misterio más allá de los límites de los sentidos— no es ni espacial ni temporal. Esto parece ofrecernos una distinción triple muy clara: lo físico es espacio-temporal; lo psicológico es no-espacial pero temporal; lo metafísico no es ni espacial ni temporal. Podemos así explicar la aparente sinonimia de «físico» y «material» como una confusión entre «no-psicológico» y «no-metafísico». El único problema es que Kant y Strawson han presentado argumentos convincentes en favor de la afirmación de que sólo podemos identificar los estados mentales en cuanto estados de personas localizadas espacialmente [1]. Como hemos abandonado «lo mental», nos vemos obligados a considerar estos argumentos con toda seriedad. Esto nos hace describir un círculo casi completo, pues ahora queremos saber qué sentido tiene decir que algunos estados de una entidad espacial son espaciales y otros no lo son. No sirve de nada decir que son sus estados *funcionales* —pues la belleza de una persona, y su figura, y su fama, y su salud, son estados funcionales, y sin embargo la intuición nos dice que no son tampoco estados mentales. Para clarificar nuestra intuición, debemos identificar un rasgo compartido por nuestros dolores y creencias, pero no por nuestra belleza o nuestra salud. No servirá de nada identificar lo mental como lo que puede sobrevivir a la muerte o a la destrucción del cuerpo, pues la propia belleza puede sobrevivir a la muerte y la fama puede sobrevivir a la destrucción del cuerpo. Si decimos que la belleza o la fama de una persona sólo existen relacionalmente, a los ojos o ante la opinión de los demás, y no en cuanto estados de la propia persona, nos metemos en problemas muy complicados sobre la forma de distinguir entre las propiedades meramente relacionales de las personas y sus estados intrínsecos. Tropezamos también con problemas igualmente enrevesados sobre las creencias

[1] Véase «Refutación del Idealismo», de Kant, en K.d.r.V. B274ss., y P. F. Strawson, *Individuals* (Londres, 1959), cap. 2, y *The Bounds of Sense* (Londres, 1966), págs. 162 y ss.

inconscientes de una persona, que sólo pueden descubrirse después de su muerte gracias a los psico-biógrafos, pero que, probablemente, son sus estados mentales con el mismo derecho que las creencias que tuvo conciencia de tener durante su vida. Puede haber una forma de explicar por qué la belleza de una persona es una propiedad relacional no intrínseca mientras que su paranoia inconsciente es un estado intrínseco no relacional; pero parece que se trata de explicar lo oscuro recurriendo a lo más oscuro.

Mi conclusión es que no podemos hacer de la no-espacialidad el criterio de los estados mentales, aunque no sea más que porque la idea de «estado» es lo suficientemente oscura como para que no parezcan útiles ni el término *estado espacial* ni el término *estado no-espacial*. La idea de que las entidades mentales son no-espaciales y de que las entidades físicas son espaciales, si es que tiene algún sentido, sólo lo tiene en relación con los particulares, con los sujetos de la predicación, y no en relación con la posesión de propiedades por tales sujetos. Podemos extraer una especie de vago sentido pre-kantiano de fragmentos de materia y fragmentos de sustancia mental, pero no podemos extraer ningún sentido post-kantiano de los estados espaciales y no-espaciales de los particulares espaciales. Recibimos una vaga sensación de capacidad explicativa cuando nos dicen que los cuerpos humanos se mueven en la forma en que lo hacen porque están habitados por fantasmas, pero no nos produce ninguna sensación semejante el escuchar que las personas tienen estados no espaciales.

Espero haberme explicado lo suficiente como para hacer ver que no tenemos derecho a comenzar a hablar del problema de la dualidad mente-cuerpo, o de la posible identidad o no-identidad necesaria de los estados mentales y físicos, sin preguntarnos antes qué entendemos por «mental». Espero también haber hecho aparecer la sospecha de que lo que llamamos intuición sobre lo que es mental quizá no sea más que nuestra disposición a aceptar un determinado juego lingüístico específicamente filosófico. Esta es, de hecho, la opinión que quiero defender. Creo que esta llamada intuición no es más que la capacidad de dominar un determinado vocabulario técnico —que no tiene ninguna utilidad fuera de los libros de filosofía y que no está en contacto con ninguna de las realidades de la vida diaria, la ciencia empírica, la moral o la religión. En las posteriores secciones de este capítulo trataré de esbozar una explicación histórica de cómo apareció este vocabulario técnico, pero antes de hacerlo, voy a considerar algunas cuestiones afines. Me refiero a las posibilidades de definir lo «mental» haciendo referen-

cia a la noción de «intencionalidad» y a la de ser «fenoménico» —de tener una apariencia característica, una apariencia en cierta forma exhaustiva de la realidad.

2. Lo funcional, lo fenoménico y lo inmaterial

La objeción más obvia a la definición de lo mental en cuanto intencional es que los dolores no son intencionales —no representan nada, no se refieren a nada. La objeción más evidente a la definición de lo mental como «lo fenoménico» es que las creencias no se parecen a nada —no tienen propiedades fenoménicas, y las verdaderas creencias de una persona no son siempre lo que parecen ser. El intento de unir los dolores y creencias parece lógico —da la impresión de que no tienen nada en común, como no sea nuestra negativa a llamarlos «físicos». Es cierto que podemos hacer divisiones arbitrarias y convertir el dolor en la adquisición de una creencia de que se ha lastimado uno de los tejidos del propio organismo, interpretando los informes de dolor como Pitcher y Armstrong interpretan los informes perceptivos [2]. Pero esta táctica sigue dejándonos con algo parecido a una intuición dualística entre las manos —la intuición de que en el ser consciente de un dolor o de una sensación de verdor hay «algo más» que el sentir la tentación de adquirir una creencia de que hay un tejido dañado o un objeto verde en los alrededores. Otra posible solución sería hacer una división artificial de signo contrario y limitar el término *mental* a lo que *sí* tiene propiedades fenoménicas, dejando las creencias y deseos en manos de Armstrong para que los identifique con lo físico. Pero esta táctica tropieza con la intuición de que sea cual sea el problema mente-cuerpo, no es el problema sensación-neurona. Si expulsamos de la mente las representaciones, los estados intencionales nos quedamos con algo que se parece al problema de la relación entre la vida y la no-vida, más que al problema mente-cuerpo.

Otra posible táctica sería sencillamente definir lo «mental» en forma disyuntiva, es decir como «o fenoménico o intencional». Esta sugerencia no explica en absoluto cómo se introdujo en el lenguaje. o al menos en la jerga filosófica, una abreviatura de esta disyunción. Sin embargo, dirige nuestra atención hacia la posibilidad de que

[2] Véase George Pitcher, *A Theory of Perception* (Princeton, 1971); D. M. Armstrong, *Perception and the Physical World* (Londres y Nueva York, 1961) y *A Materialist Theory of the Mind* (Londres y Nueva York, 1968).

los parecidos de familia mantengan unidos a ítems «mentales» diversos. Si consideramos los pensamientos —los pensamientos que ocurren y emiten destellos ante la mente en palabras concretas— o las imágenes mentales, parece que tenemos algo que se parece un poco al dolor por ser fenoménico y otro poco a la creencia por ser intencional. Las palabras hacen que los pensamientos sean fenoménicos y los colores y formas hacen que lo sean las imágenes, sin embargo, ambos son *de* algo en el sentido intencional requerido. Si, de repente, me digo a mí mismo en silencio: «Dios mío, me he dejado la cartera en aquella mesa del café de Viena», o si tengo una imagen de la cartera encima de la mesa, entonces estoy representando Viena, la cartera, la mesa, etc. —los tengo todos en cuanto objetos intencionales. Por eso, quizá deberíamos pensar en los pensamientos e imágenes mentales en cuanto entidades mentales *paradigmáticas*. Entonces podemos decir que los dolores y creencias se clasifican como mentales debido a su parecido con estos paradigmas, aun cuando dicho parecido se dé en dos aspectos muy diferentes. La relación entre los diversos aspirantes a la categoría de mental podría quedar ilustrada en el siguiente diagrama:

	con propiedades fenoménicas	fenoménicas sin propiedades
intencional, representacional	pensamientos que ocurren, imágenes mentales	creencias, deseos, intenciones
no-intencional, no-representacional	sensaciones primarias, por ejemplo, los dolores y lo que tienen los bebés cuando ven objetos de colores	«lo meramente físico»

Supongamos, de momento, que nos contentamos con esta respuesta del «parecido de familia» a la pregunta «¿qué hace que lo mental sea mental?» —es decir, que es uno u otro parecido de familia a lo paradigmáticamente mental. Volvamos ahora a la pregunta original, y preguntemos qué es lo que nos hace rellenar la cuarta casilla con «lo meramente físico». ¿Significa «físico» sencillamente «lo que no encaja en las otras tres casillas»? ¿Se trata de una noción que depende totalmente de la de «mental»? ¿O se rela-

ciona de alguna manera con «material» y «espacial», y cómo lo hace?

Para responder a esto, debemos formularnos dos sub-preguntas: «¿Por qué lo intencional es no-material?» y «¿Por qué lo fenoménico es no-material?» La respuesta a la primera pregunta parece sencilla. Si consideramos que «lo material» es «lo neural», por ejemplo, podemos decir que por mucho que se examine el cerebro no se revelará el carácter intencional de las imágenes e inscripciones que se encuentren en él. Supongamos que todas las personas a quienes se les ocurre el pensamiento «I left my wallet on a cafe-table in Vienna» [Me dejé la cartera en la mesa de un café de Viena], *con esas mismas palabras inglesas,* tienen una serie idéntica de hechos neurales concomitantes del pensamiento. Parece una hipótesis plausible (aunque probablemente falsa). Pero *no* es plausible que todos los que adquieren la creencia de que se han dejado la cartera en la mesa de un café de Viena tenga esta serie de hechos, pues pueden formular su creencia en otras palabras completamente distintas o en otra lengua diferente. Sería extraño que un pensamiento en japonés y en inglés tuviera el mismo correlativo neural. Es igualmente plausible que todos los que vean la misma cartera perdida en la misma mesa lejana en el ojo de su mente compartan una segunda serie de hechos neurales aunque muy diferente de la correlacionada con el pensamiento de la oración inglesa. Ni siquiera una concomitancia tan clara nos induciría a «identificar» las propiedades intencionales y neurológicas del pensamiento o la imagen, por lo mismo que no identificamos las propiedades tipográficas e intencionales de la oración «I left my wallet on a cafe-table in Vienna» cuando la vemos en una página impresa. Una vez más, la concomitancia de las imágenes de carteras en mesas de café en un trasfondo vienés con ciertas propiedades de las superficies del papel y el lienzo no identifica la propiedad intencional de «tratar sobre Viena» con la distribución de los pigmentos en el espacio. Por eso, podemos ver por qué se puede decir que las propiedades intencionales no son propiedades físicas. Pero por otra parte, esta comparación entre propiedades neurológicas y tipográficas indica que no hay ningún problema interesante en relación con la intencionalidad. Nadie desea buscar complicaciones filosóficas por el hecho de que no se pueda decir qué significa una oración atendiendo sólo a su apariencia, o de que no se pueda reconocer una imagen de X *en cuanto* imagen de X sin estar familiarizado con las convenciones gráficas relevantes. Parece perfectamente claro, al menos desde Wittgenstein y Sellars, que el «significado» de las inscripciones tipo-

gráficas no es una propiedad «inmaterial» que tengan por encima, sino sencillamente el lugar que ocupan en un contexto de hechos circundantes dentro de un juego lingüístico o de una forma de vida. Este se aplica también a las inscripciones cerebrales. Decir que no podemos observar las propiedades intencionales mirando al cerebro es igual que decir que no podemos ver una proposición cuando vemos un código maya —sencillamente, no sabemos qué buscar, pues todavía no sabemos cómo relacionar lo que vemos con un sistema de símbolos. La relación entre una inscripción —sobre el papel o, dada la concomitancia a que se refiere la hipótesis, en el cerebro— y lo que significa no es más misteriosa que la relación entre un estado funcional de una persona, como su belleza o salud, y los miembros de su cuerpo. Se trata precisamente de esos miembros vistos en un contexto dado.

Por eso, la respuesta a la pregunta «¿Por qué lo intencional es no-material?» es «porque todo estado funcional —todo estado que sólo pueda entenderse relacionando lo que se observa con un contexto más amplio— es, en un sentido superficial, no-material». El problema está en tratar de relacionar esta noción superficial de ser «no-material» —que significa simplemente algo parecido a «no evidente de forma inmediata para todos los que miran»— con el sentido de «inmaterialidad» que está cargado de sentido filosófico. Dicho de otra manera, ¿por qué debe preocuparnos la afirmación de Leibniz de que si se ampliara el cerebro hasta darle el tamaño de una fábrica, de forma que pudiéramos pasear por su interior, no veríamos los pensamientos? Si conocemos el número suficiente de correlaciones neurales, veremos realmente los pensamientos —en el sentido de que nuestra visión nos revelará qué pensamiento tiene el propietario de ese cerebro. Si carecemos de ese conocimiento, nuestra visión no nos revelará nada semejante, pero si recorremos una fábrica *cualquiera* sin habernos informado previamente de sus elementos y de sus mutuas relaciones, no veremos qué es lo que ocurre. Además, aun cuando no pudiéramos encontrar estas correlaciones neurales, aun cuando la localización cerebral de los pensamientos fuera un completo fracaso, ¿por qué vamos a decir que los pensamientos o imágenes mentales de una persona eran no-físicos sencillamente porque no podemos explicarlos haciendo referencia a sus elementos? Utilizando un ejemplo de Hilary Putnam, no podemos presentar una explicación de por qué las clavijas cuadradas no encajan en los agujeros redondos haciendo referencia a las partículas elementales que constituyen la clavija y el agujero, pero nadie

considera que se dé un sorprendente abismo ontológico entre macroestructura y microestructura.

Creo que sólo podemos relacionar el sentido superficial de «no-material» (que se aplica a cualquier estado funcional, en cuanto opuesto a observable) con el sentido rico en contenido de «inmaterial» resucitando el punto de vista de Locke sobre la forma en que se relaciona el significado con las inscripciones —el punto de vista que atacan Wittgenstein y Sellars. Para Locke, el carácter significativo —el carácter intencional— de una inscripción era resultado de su producción por, o codificación de, una idea. A su vez, una idea era «lo que está ante la mente de un hombre cuando piensa». Por eso, la forma de ver lo intencional como inmaterial es decir que ni una secuencia de procesos del cerebro ni la tinta que hay sobre un papel pueden representar algo a no ser que las haya impregnado una idea, algo de lo que somos conscientes en la forma «inmediata» en que somos conscientes de los dolores. En una visión lockeana, cuando recorremos la fábrica de que habla Leibniz no vemos pensamientos no porque, como para Wittgenstein, no sepamos todavía traducir la escritura cerebral, sino porque no podemos ver las entidades invisibles (por no-espaciales) que infunden intencionalidad en lo visible. Para Wittgenstein, lo que hace que las cosas sean representacionales o intencionales es el papel que desempeñan en un contexto más amplio —en interacción con gran número de otras cosas visibles. Para Locke, lo que hace que las cosas sean representacionales es un empuje causal especial —lo que Chisholm describe como el fenómeno de las oraciones que extraen su intencionalidad de los pensamientos, de forma semejante a como la luna recibe su luz del sol [3].

Por eso, nuestra respuesta a la pregunta «¿Cómo podemos convencernos de que lo intencional debe ser inmaterial?» es: «Primero debemos convencernos, siguiendo a Locke y a Chisholm, y a pesar de Wittgenstein y Sellars, de que la intencionalidad es intrínseca únicamente en los aspectos fenoménicos —aspectos que están directamente ante la mente.» Sin embargo, si aceptamos esta respuesta, sólo hemos hecho parte del recorrido que puede llevarnos a la solución de la dificultad. Como el problema a que nos venimos enfrentando ha sido producido precisamente por el hecho de que las creencias no tienen propiedades fenoménicas, ahora debemos preguntarnos cómo puede Locke, siguiendo a Descartes, incluir los

[3] Roderick Chisholm, «Intentionality and the Mental» en *Minnesota Studies in the Philosophy of Science,* 2 (1958), 533.

dolores y creencias bajo el término común *idea,* cómo puede convencerse de que una creencia es algo que está «delante de la mente» en la forma en que lo está una imagen mental, cómo puede utilizar el mismo conjunto de imágenes oculares para las imágenes mentales y para los juicios. Más adelante examinaré el origen de este uso cartesiano-lockeano del término. Pero, de momento, voy a dejar de lado la cuestión y consideraré la *segunda* subdivisión de la pregunta «¿Por qué debe considerarse que lo mental es inmaterial?» —es decir, ¿por qué debe considerarse que lo *fenoménico* es inmaterial? ¿Por qué dicen algunos filósofos neo-dualistas que el cómo se encuentra alguien, o a qué se parece ser algo, no pueden ser idénticos a ninguna propiedad física, o al menos a ninguna propiedad física de la que sepamos algo?

Una respuesta superficial a esta cuestión sería decir que podemos saberlo todo sobre las propiedades físicas de algo y no saber cómo se encuentra —especialmente si no podemos hablar con el ser en cuestión. Consideremos la afirmación de que los niños y los murciélagos y los marcianos y Dios y las rocas consideradas desde un punto de vista panpsíquico pueden habitar «espacios cualitativos» fenoménicos diferentes de los que nosotros habitamos [4]. Me alegro por ellos. Pero, ¿qué tiene que ver esto con lo no-físico? Probablemente, los que dicen que lo fenoménico es no-físico no se lamentan de que el saber cómo están dispuestos los átomos del cerebro del murciélago no sirva de ayuda para sentirse como uno

[4] Esta afirmación aparece presentada con mucha fuerza en Thomas Nagel, «What Is It Like to Be a Bat?», *Philosophical Review,* 83 (1974), 435-450. He aprendido mucho en las obras de Nagel sobre filosofía de la mente, aunque estoy en profundo desacuerdo con él casi en todos los puntos. Creo que la diferencia entre nuestros puntos de vista se remonta a la cuestión (agudamente suscitada por Wittgenstein) de si las «intuiciones filosóficas» son algo más que residuos de las prácticas lingüísticas, pero no estoy seguro de cómo se debe debatir este punto. La intuición de Nagel le dice que «los hechos sobre qué es probable que sea X son muy peculiares» (pág. 437), mientras que yo estimo que parecen peculiares únicamente si, siguiendo la tradición cartesiana y a Nagel, afirmamos que «para no tener que defender el fisicismo, hay que presentar una explicación física de los mismos rasgos fenomenológicos» (pág. 437). En secciones posteriores de este capítulo trato de seguir la historia del juego lingüístico-filosófico en que esta afirmación parece sentirse como en casa. Por las razones davidsonianas presentadas en el capítulo cuarto, sección 4, más adelante, no creo que el fisicismo esté sometido a semejante constricción. El fisicismo, afirmo en dicho lugar, es probablemente cierto (pero carente de interés) si se interpreta como predicción de todos y cada uno de los hechos en todas y cada una de las regiones del espacio-tiempo bajo una u otra descripción, pero evidentemente falso si se interpreta como la pretensión de decir que todo es verdad.

de ellos. Tampoco la comprensión de la fisiología del dolor, pero ¿por qué voy a esperar que lo haga en mayor medida que la comprensión de la aerodinámica nos pueda ayudar a volar? ¿Cómo del hecho indubitable de que el saber utilizar un término fisiológico (por ejemplo, «estimulación de las fibras C») no nos ayudará necesariamente a utilizar un término fenomenológico (por ejemplo, «dolor») podemos pasar a la existencia de un abismo ontológico entre los referentes de los dos términos? ¿Cómo del hecho de que conocer la fisiología de los marcianos no nos sirve para traducir lo que diga un marciano cuando le lesionamos los tejidos podemos concluir la afirmación de que tiene algo inmaterial que nosotros no tenemos? ¿Cómo, centrándonos en nuestro tema, sabemos cuándo tenemos dos formas de hablar de la misma cosa (una persona, o su cerebro) y no descripciones de dos cosas diferentes? ¿Y por qué están tan seguros los neo-dualistas de que los sentimientos y neuronas son un caso de esto último?

Creo que la única respuesta que pueden presentar estos filósofos es señalar que en el caso de las propiedades fenoménicas no hay ninguna distinción apariencia-realidad. Esto equivale a definir una propiedad física como aquella que cualquiera pudiera atribuir erróneamente a algo, y propiedad fenoménica aquella sobre la que una determinada persona no se puede equivocar. (Por ejemplo, la persona que tiene el dolor no puede equivocarse sobre la sensación que produce el dolor). Dada esta definición, lo lógico es concluir que ninguna propiedad fenoménica puede ser una propiedad física. Pero, ¿por qué debe reflejar esta distinción *epistémica* una distinción *ontológica?* ¿Por qué el privilegio epistémico que tenemos todos de que nadie pueda corregirnos al hablar de cómo nos parecen las cosas debe reflejar una distinción entre dos campos del ser?

La respuesta deberá estar en esta línea: Las sensaciones *son* sólo apariencias. Su realidad se agota en lo que parecen. Son puros pareceres. Todo lo que no es apariencia (dejando por el momento de lado lo intencional) es meramente físico —es decir, es algo que puede aparecer de distinta manera a como es. El mundo está dividido en cosas cuya naturaleza se agota en su apariencia y cosas que no. Pero si un filósofo presenta esta respuesta corre el peligro de dejar de ser un neodualista para convertirse sencilla y llanamente en un dualista cartesiano al viejo estilo, con la «sustancia mental» y todo. Pues en este momento ha dejado de hablar sobre los dolores en cuanto estados de las personas o propiedades predicadas de las personas, y ha comenzado a hablar sobre los dolores en cuanto particulares, una clase especial de particulares cuya naturaleza se

agota en una sola propiedad. ¿De qué podría estar hecho este particular, como no sea de sustancia mental? O, dicho de otra manera, ¿qué podría ser la sustancia mental como no sea algo de lo que se pueden hacer cosas ligeras, finas y translúcidas? Mientras el sentir dolor sea una propiedad de la persona o de las fibras cerebrales, parece que no hay ninguna razón que justifique que la diferencia epistémica entre los informes de cómo se sienten las cosas y los informes de cualquier otra cosa produzcan un abismo ontológico. Pero en el mismo momento en que haya un abismo ontológico, ya no estamos hablando sobre los estados o propiedades sino sobre particulares distintos, sobre diferentes sujetos de predicación. El neo-dualista que identifica un dolor con la sensación de tener dolor está hipostatizando una propiedad —la propiedad de ser doloroso— y convirtiéndola en una especie de particular, un particular de esa clase especial cuyo *esse* es *percipi* y cuya realidad se agota en nuestro contacto inicial con él. El neo-dualista ya no habla sobre cómo se encuentran las personas sino sobre las sensaciones en cuanto pequeñas entidades auto-subsistentes, que flotan independientemente de las personas de la misma manera que los universales flotan con independencia de las instanciaciones. De hecho, ha reproducido en los dolores el modelo de los universales. No es de extrañar por lo tanto, que pueda «intuir» que los dolores pueden existir separados del cuerpo, pues esta intuición es sencillamente la intuición de que los universales pueden existir independientemente de los particulares. Esa clase especial de sujeto de predicación cuya apariencia *es* su realidad —dolor fenoménico— resulta ser sencillamente el carácter doloroso del dolor abstraído de la persona que lo tiene. Es, en resumen, el mismo universal de *lo doloroso*. Dicho en forma más llamativa, los particulares mentales, a diferencia de los estados mentales de las personas, resulta que son universales.

Esta es, pues, la respuesta que yo quiero dar a la pregunta: ¿Por qué pensamos que lo fenoménico es inmaterial? Lo hacemos porque, como señaló Ryle, insistimos en pensar en lo que es tener un dolor con metáforas oculares —como tener una forma curiosa de particular ante el ojo de la mente. Ese particular resulta ser un universal, una cualidad hipostatizada en un sujeto de predicación. Así, cuando los neo-dualistas dicen que la forma en que se *sienten* los dolores resulta esencial para lo que *son,* y luego critican a Smart por pensar en el papel causal de ciertas neuronas como lo que es esencial para el dolor, están cambiando de tema. Smart está hablando de lo que es esencial para que las personas tengan dolor, mientras que los neo-dualistas como Kripke están

hablando de lo que es esencial para que algo sea *un dolor*. Los neodualistas no se asustan ante la pregunta «¿Cuál es la base epistemológica para su pretensión de saber qué es una propiedad esencial del dolor?», pues han dispuesto las cosas de tal manera que los dolores sólo tengan *una* propiedad intrínseca —a saber, la sensación dolorosa— y por tanto la elección de cuáles son las propiedades que deben figurar como esenciales resulta evidente.

Resumamos los resultados de esta sección. He dicho que la única manera de asociar lo intencional con lo inmaterial es identificarlo con lo fenoménico, y que la única forma de identificar lo fenoménico con lo inmaterial es hipostatizar los universales y pensar en ellos en cuanto particulares más que como abstracciones de los particulares —dándoles así una morada no-espacio-temporal. En otras palabras, resulta que la distinción universal-particular es la *única* distinción metafísica con que contamos, la única que llega a moverse algo fuera del espacio, y mucho menos fuera del espacio-tiempo. La distinción mental-físico es, por tanto, dependiente de la distinción universal-particular, y no al revés. Además, la noción de la sustancia mental como aquello de lo que están hechos los dolores y creencias tiene exactamente tanto o tan poco sentido como la noción de «aquello de que están hechos los universales». La batalla entre realistas y conceptualistas por la categoría de los universales queda así vacía de contenido, pues no tenemos la más mínima idea de qué es la mente, a no ser que está hecha de lo mismo que estén hechos los universales. Al construir una idea lockeana y una Forma platónica seguimos exactamente el mismo proceso —sencillamente, extraemos una sola propiedad de algo (la propiedad de ser rojo, o doloroso, o bueno) y luego la tratamos como si fuera ella misma un sujeto de predicación, y quizá también un lugar de eficacia causal. Una Forma platónica es simplemente una propiedad considerada de forma aislada y como si fuera capaz de mantener relaciones causales. Una entidad fenoménica es también eso mismo.

3. La diversidad de los problemas mente-cuerpo

Al llegar a este punto podríamos intentar decir que hemos acabado con el problema mente-cuerpo. Efectivamente, y hablando en términos generales, todo lo que hace falta para considerar que este problema es ininteligible es que seamos nominalistas, que nos neguemos firmemente a hipostatizar las propiedades individua-

les. Entonces no nos dejaremos engañar por la idea de que hay entidades llamadas dolores que, por ser fenoménicas, no pueden ser físicas. Siguiendo a Wittgenstein, consideraremos el hecho de que no haya una cosa semejante a «una apariencia engañosa de dolor» no como un hecho extraño relacionado con un género ontológico especial llamado mental, sino simplemente como un comentario sobre un juego lingüístico —el comentario de que tenemos el acuerdo de aceptar lo que los otros nos dicen sobre cómo se encuentran. Desde este punto de vista del «juego lingüístico», el hecho de que un hombre sienta cualquier cosa que crea que siente no tiene más importancia ontológica que el hecho de que la Constitución sea lo que considera que es el Tribunal Supremo, o que la pelota ha salido fuera si el árbitro así lo considera. Siguiendo también a Wittgenstein, trataremos lo intencional meramente como una subespecie de lo funcional, y lo funcional meramente como la clase de propiedad cuya atribución depende de un conocimiento del contexto y no como algo que sea observable de forma inmediata. Veremos que lo intencional no tiene ninguna conexión con lo fenoménico, y que lo fenoménico es una cuestión relacionada con nuestra forma de hablar. El problema mente-cuerpo, podemos decir ahora, fue simplemente resultado de un error desafortunado de Locke sobre la forma en que las palabras reciben significado, junto con el intento suyo y de Platón de hablar sobre los adjetivos como si fueran nombres.

Para lo que suelen ser las eliminaciones de los problemas filosóficos, ésta tiene sus puntos de interés. Pero sería absurdo pensar que habíamos resuelto algo por el mero hecho de llegar a este diagnóstico. Es como si un psiquiatra le explicara al paciente que su infelicidad es consecuencia de su creencia errónea de que su madre había querido castrarle, junto con su confuso intento de pensar en sí mismo como si fuera igual que su padre. Lo que el paciente necesita no es una lista de sus errores y confusiones sino más bien una comprensión de cómo llegó a cometer estos errores y a verse inmerso en estas confusiones. Si queremos librarnos del problema mente-cuerpo, debemos ser capaces de responder a preguntas como éstas:

> ¿Cómo llegaron las cuestiones vagas y diminutas relacionadas con la posible identidad de los dolores y las neuronas a mezclarse con la cuestión de si el hombre era «de una clase diferente» a los brutos —si tenía dignidad y no meramente un valor?

> Dado que las personas pensaban que sobrevivían a la destrucción de sus cuerpos mucho antes de que Locke y Platón comenzaran a crear confusiones específicamente filosóficas, ¿no hemos dejado algo fuera cuando tratamos la mente como si fuera simplemente una acumulación de estados fenoménicos e intencionales?
>
> ¿No existe alguna conexión entre nuestra capacidad de tener conocimiento y el que tengamos una mente, y se explica esto haciendo referencia simplemente al hecho de que las personas, como las inscripciones, tienen propiedades intencionales?

Todas ellas son preguntas bien planteadas, y nada de lo que he dicho hasta ahora sirve para darles respuesta. Para ello, pienso yo, lo único que servirá es la historia de las ideas. Lo mismo que el paciente necesita revivir su pasado para responder a sus preguntas, también la filosofía necesita revivir su pasado para responder a *sus* preguntas. En las páginas anteriores, tal como es habitual en los actuales filósofos de la mente he usado a diestro y siniestro términos como «fenoménico», «funcional», «intencional», «espacial» y otros semejantes como si formaran el vocabulario evidente en que se debe estudiar el tema. Pero, lógicamente, los filósofos que crearon el lenguaje que nos dio el problema mente-cuerpo no utilizaron este vocabulario, ni nada que se le pareciera. Si queremos entender cómo conseguimos las intuiciones que nos hicieron pensar que *tiene* que haber *por alguna parte* un problema filosófico indisoluble *real,* debemos dejar de lado nuestra jerga actualizada y pensar en el vocabulario de los filósofos cuyos libros nos ofrecieron esas intuiciones. Desde mi punto de vista wittgensteiniano, una intuición no es nunca más o menos que la familiaridad con un juego lingüístico, por lo que descubrir la fuente de nuestras intuiciones es revivir la historia del juego lingüístico filosófico que estamos jugando.

El «problema mente-cuerpo» que acabo de «suprimir» se refiere únicamente a unas pocas de las nociones que, surgidas en distintos momentos de la historia del pensamiento, han llegado a entrelazarse, produciendo una maraña de problemas interrelacionados. Preguntas del tipo «¿Cómo se relacionan los estados intencionales de la conciencia con los estados neurales?» y «¿Cómo se relacionan propiedades fenoménicas como el dolor con las propiedades neurológicas?» forman parte de lo que voy a llamar el «problema de la conciencia». Este problema es distinto de algunos problemas pre-filosóficos sobre la persona, como pueden ser: «¿Cómo podemos tener seguridad sobre el cambio?» «¿Cómo pue-

de haber conocimiento de lo que no cambia?» y «¿Cómo puede lo que no cambia llegar a ser interior a nosotros por el hecho de ser conocido?» Vamos a llamar «problema de la personalidad» al que se refiere a qué es el sujeto humano además de carne. Este problema tiene una forma en la aspiración pre-filosófica a la inmortalidad, y otra en la afirmación kantiana y romántica de la dignidad humana —pero ambos deseos son muy distintos de los problemas sobre la conciencia y sobre el conocimiento. Ambos son formas de expresar nuestra pretensión de ser algo completamente diferente de las bestias perecederas. Llamamos «problema de la razón» al de explicar la afirmación griega de que la diferencia fundamental entre hombres y bestias es que nosotros podemos *conocer* —que podemos conocer no simplemente los hechos singulares sino las verdades universales, los números, las esencias, lo eterno. Este problema adopta formas diferentes en la explicación hilemórfica del conocimiento presentada por Aristóteles, en la explicación racionalista de Spinoza o en la explicación transcendental de Kant. Pero estas cuestiones se distinguen de las que tratan de las interrelaciones entre dos clases de cosas (una espacial y otra no-espacial) y de las que se refieren a la inmortalidad y a la dignidad moral. El problema de la conciencia se centra en torno al cerebro, las sensaciones primarias, y los movimientos corporales. El problema de la razón gira en torno a los temas del conocimiento, el lenguaje y la inteligencia —todas nuestras «potencias superiores». El problema de la personalidad se centra en las atribuciones de libertad y responsabilidad moral.

Para ordenar algunas de las relaciones existentes entre estos tres problemas, voy a presentar una enumeración de las formas de aislar seres que tengan mentes en contraste con lo «meramente físico» —«el cuerpo», «la materia», el sistema nervioso central, la «naturaleza» o «el objeto material de las ciencias positivas». He aquí algunos, pero no todos, de los rasgos que, en uno u otro momento, los filósofos han considerado como marcas de lo mental:

1. capacidad de conocerse a sí mismo incorregiblemente («acceso privilegiado»)
2. capacidad de existir independientemente del cuerpo
3. no-espacialidad (tener una parte o «elemento» no-espacial)
4. capacidad de captar los universales
5. capacidad de mantener relaciones con lo inexistente («intencionalidad»)
6. capacidad de usar el lenguaje
7. capacidad de actuar libremente

8. capacidad de formar parte de nuestro grupo social, de ser «uno de nosotros»
9. incapacidad de ser identificado con cualquier objeto «del mundo»

La enumeración es larga, y sería muy fácil continuar [5]. Pero es importante recorrer estas distintas sugerencias sobre lo que significa tener mente, pues cada una de ellas ha ayudado a los filósofos a insistir en que hay un dualismo insalvable entre la mente y el cuerpo. Los filósofos han echado siempre mano de algún rasgo distintivo de la vida humana para dar una «firme base filosófica) a la intuición de nuestra singularidad. Como estas bases firmes son tan variadas, los naturalismos y materialismos, cuando no quedan reducidos a meros intentos desesperados de saltar un enorme abismo ontológico (o epistemológico o lingüístico), suelen ser tratados como superficialmente verdaderos pero carentes de sentido. No tienen sentido, se explica, porque nuestra singularidad no tiene nada que ver con el abismo, cualquiera que sea, que haya tratado de colmar el naturalista, y sí tienen mucho que ver con algún otro abismo que ha tenido abierto en todo momento, justo a sus espaldas. En concreto, se señala muchas veces que aun cuando solucionáramos todas las cuestiones sobre la relación entre dolores y neuronas, y las cuestiones semejantes derivadas de la incorregibilidad —véase (1), más arriba— sólo habríamos solucionado, en el mejor de los casos (2) y (3), entre las demás señales de lo mental. Todo lo que hace referencia a al razón (especialmente [4], [5] y [6]) y todo lo que se refiere a la personalidad (sobre todo [7], [8] y [9]) seguiría estando todavía tan oscuro como antes.

Creo que esta precisión está totalmente en lo cierto, y, además, que si se hubiera visto antes, el problema de la conciencia no habría adquirido la importancia que ha tenido en la filosofía reciente. En el sentido de tener tanto dolores como neuronas, estamos en situación de igualdad con muchas, si no todas, las bestias, mientras que es de suponer que no compartimos con ellas ni la razón ni la personalidad. Sólo si suponemos que la posesión de *cualquier* estado interior no-físico está relacionada de alguna manera, a través de (3), con (4) o (5), pensaremos que la luz arrojada sobre las sensaciones primarias se reflejaría sobre los estados mentales representacionales, y por lo tanto iluminaría nuestra capaci-

[5] Véase Herbert Feigl, *The «Mental» and the «Physical»* (Minneapolis, 1967), donde se presenta una enumeración semejante y comentarios reveladores sobre las relaciones entre los varios puntos.

dad de reflejar el mundo que nos rodea. De la misma manera, sólo la suposición de que la vida misma (incluso la del feto, la del ser humano con lesiones cerebrales, la del murciélago o la de la oruga) tiene un carácter sagrado especial semejante a la personalidad, nos haría pensar que la comprensión de las sensaciones primarias podría ayudarnos a comprender nuestras responsabilidades morales. Sin embargo, ambas suposiciones se hacen con mucha frecuencia. Para comprender por qué se hacen, es necesario comprender la historia intelectual más que entender los significados de los términos relevantes o un análisis de los conceptos que significan. Al esbozar una pequeña parte de la historia de las discusiones sobre la mente, espero hacer ver que el problema de la razón no se puede formular sin una vuelta a concepciones epistemológicas que nadie quiere realmente resucitar. Además, quiero presentar algunas pruebas en favor de la sugerencia que desarrollaré más adelante: que el problema de la personalidad no es un «problema» sino una descripción de la condición humana, es decir que no es un tema que deba aspirar a una «solución» filosófica sino una forma engañosa de protestar por la irrelevancia de la filosofía tradicional para el resto de la cultura.

Sin embargo, no me voy a ocupar de todos los puntos mencionados anteriormente, sino únicamente del (2), el (3) y el (4) —separación del cuerpo, no-espacialidad y comprensión de los universales. Lo que quiero decir sobre los demás puntos irá apareciendo en los otros capítulos. En el siguiente, examinaré (1) —el acceso privilegiado—, y en los capítulos cuatro y seis me ocuparé de (5) y (6) —la intencionalidad y la capacidad de utilizar el lenguaje. Aunque los aspectos que se refieren a la personalidad —(7), (8) y (9)— no van a ser examinados por separado, señalaré brevemente cómo creo que debe tratarse la noción de personalidad (sección 4 del capítulo cuatro, sección 4 del capítulo siete y sección 3 del capítulo ocho). En el presente capítulo quiero ceñirme lo más posible a la pregunta: ¿Por qué ha de parecer que la conciencia tiene algo que ver con la razón o con la personalidad? Fijándome en los tres temas de la captación de los universales, la separación del cuerpo y la no-espacialidad, avanzaré hacia mi conclusión de que si mantenemos apartadas estas tres nociones históricamente distintas, ya no nos sentiremos tentados por más tiempo por la noción de que el conocimiento es posible gracias a una Esencia de Vidrio especial que permite a los seres humanos reflejar la naturaleza. De esta manera no sentiremos la tentación de pensar que la posesión de una vida interior, una corriente de la conciencia,

tiene relación con la razón. Una vez que se separan de esta manera la conciencia y la razón, se verá la personalidad tal como yo digo: como una cuestión de decisión, más que de conocimiento, una aceptación de otro ser para formar una visión, y no un reconocimiento de una esencia común.

4. La mente en cuanto comprensión de los universales

No se habría llegado a pensar que había un problema con la naturaleza de la razón si nuestra raza se hubiera limitado a señalar estados de cosas particulares —prevenir de los acantilados y de la lluvia, celebrar los nacimientos y muertes individuales. Pero la poesía habla del hombre, el nacimiento y la muerte en cuanto tales, y la matemática se precia de pasar por alto los detalles individuales. Cuando la poesía y las matemáticas hubieron llegado a la auto-conciencia —cuando hombres como Ión y Teeteto podían identificarse con sus sujetos— llegó la hora de decir algo general sobre el conocimiento de los universales. La filosofía se dedicó a examinar la diferencia entre saber que había cadenas montañosas paralelas que se extendían hacia el oeste y saber que las líneas paralelas no se cruzan aunque se extiendan hasta el infinito, la diferencia entre saber que Sócrates era bueno y saber qué era la bondad. Así surgió la pregunta: ¿Cuáles son las analogías entre saber de montañas y saber de líneas, entre saber de Sócrates y saber del Bien? Cuando esta cuestión recibió respuesta en términos de la distinción entre el ojo del cuerpo y el Ojo de la Mente, el νοῦς —pensamiento, intelecto, perspicacia— fue identificado como aquello que separa a los hombres de las bestias. No había ninguna razón, podemos decir los hombres posteriores con la ingratitud de la retrovisión, para que esta metáfora ocular se apoderara de la imaginación de los fundadores del pensamiento occidental. Pero lo hizo, y los filósofos de nuestro tiempo están todavía sacando sus consecuencias, analizando los problemas que creó, y preguntándose si, después de todo, no habrá algo de verdad en ello. La idea de la «contemplación», del conocimiento de los conceptos o verdades universales en cuanto θεωρία, convierte al Ojo de la Mente en el modelo inevitable de la forma superior de conocimiento. Pero es inútil preguntarse si la lengua griega, o las condiciones económicas griegas, o la imaginación ociosa de algún pre-socrático innominado, son responsables de que se considere este tipo de conocimiento como *mirar* algo (en vez de, por ejemplo, restregarse con ello, o aplastarse

con ello, o aplastarlo bajo el pie, o tener relaciones sexuales con ello)[6].

Dado este modelo, y con él el Ojo de la Mente, ¿qué debe ser la mente? Podemos suponer que algo tan diferente del cuerpo como lo es el paralelismo invisible de las cadenas montañosas visibles. Podía echarse mano de algo parecido a esto, pues la poesía y la religión indicaban que al morir una realidad humanoide deja el cuerpo y sigue viviendo por su cuenta[7]. El paralelismo puede considerarse como el aliento mismo de las paralelas —la sombra que queda cuando ya no hay montañas. Cuanto más fina sea la

---

[6] Dewey considera la metáfora del Ojo de la Mente como resultado de la noción previa de que el conocimiento ha de ser inmutable:

> La teoría del conocimiento se inspira en el modelo de lo que se supone que ocurre en el acto de visión. El objeto refracta la luz y es visto; si se tiene un instrumento óptico pueden notar un cambio el ojo y la persona, pero no la cosa vista. El objeto real es el objeto tan fijo en su distanciamiento regio que es un rey para cualquier mente que pueda poner en él la mirada. El resultado inevitable es una teoría espectadora del conocimiento. *(The Quest for Certainty* [Nueva York, 1960], pág. 23.)

Es difícil saber si la metáfora óptica determinó la idea de que el objeto del conocimiento verdadero ha de ser eterno e inmutable o al revés, pero parece que las dos ideas están hechas la una para la otra. Compárese con A. O. Lovejoy, *The Great Chain of Being* (Cambridge, Mass., 1936), cap. 2. No obstante, la búsqueda de certeza y la metáfora óptica persisten una vez que se abandona la idea de inmutabilidad y eternidad —por ejemplo, el argumento de C. D. Broad sobre los datos sensoriales basándose en que «si no hay nada elíptico ante la mente resulta muy difícil entender por qué el penique ha de parecer elíptico y no de ninguna otra forma» *(Scientific Thought* [Londres, 1923], página 240).

[7] Sobre la conexión entre ψυχή, sombra y aliento, véase C. A. van Peursen, *Body, Soul, Spirit* (Oxford, 1966), pág. 88 y cap. 7 *passim,* junto con los pasajes de Bruno Snell, *Discovery of the Mind* (Cambridge, Mass., 1953) y R. B. Onians, *The Origins of European Thought* (Cambridge, Mass., 1951) a que se refiere Peursen. Onians, al examinar la relación entre θυμός y ψυχή (págs. 93 y ss.), aclara la poca conexión que tenían ambas nociones con el conocer y su intensa relación con el combate, el sexo y el movimiento en general. Sobre la relación de estas dos ideas con el νοῦς en el periodo prefilosófico, véase Snell, cap. I, donde se cita y explica la descripción que hace Platón del νοῦς en cuanto «ojo del alma» y se hace referencia al uso arcaico de νοεῖν en cuanto captación de imágenes. En el plano en que nos estamos moviendo, lo importante es que la distinción clara entre, como dice van Peursen, «mundos interiores y exteriores» sólo se llega a desarrollar cuando se da la idea de un objeto de conocimiento inmaterial e invisible (como en el conocimiento del geómetra). Cfr. van Peursen, págs. 87 y 90.

mente, más capacitada estará para divisar entidades invisibles como el paralelismo. Por eso, hasta Aristóteles, que dedicó su vida a arrojar agua fría sobre las extravagancias metafísicas de sus predecesores, sugiere que es probable que haya *algo* de cierto en la idea de que el intelecto es «separable», aunque sea lo único que se pueda decir sobre el alma. Los seguidores de Ryle y Dewey han alabado a Aristóteles para haber hecho frente al dualismo al pensar que el «alma» no era ontológicamente más distinta del cuerpo de lo que lo eran las capacidades de la rana para cazar moscas y huir de las serpientes y el cuerpo de la misma rana. Pero esta visión «naturalista» del alma no impidió a Aristóteles afirmar que, puesto que el intelecto tenía la capacidad de recibir la forma de la «raneidad» (extrayendo lo universal de la rana concreta conocida, por así decirlo) y asumirla sin convertirse por ello en rana, el intelecto (νοῦς) debe ser realmente algo muy especial. Tiene que ser algo inmaterial —aun cuando no sea necesario postular esta extraña cuasi-sustancia para explicar la *mayor parte* de la actividad humana, en la misma medida en que no es necesario postularla para explicar la de la rana [8].

Los filósofos han expresado muchas veces sus deseos de que Aristóteles no hubiera accedido nunca a hablar como Platón de los universales ni hubiera aceptado su teoría auxiliar del conocimiento, o que su *Entwicklung* hubiera durado el tiempo suficiente para que se pudieran descartar como obras de juventud pasajes como *De Anima* III, 5 y *Metafísica XII* [9]. Pero, una vez

[8] Creo que Aristóteles no presentó nunca explícitamente este argumento para demostrar que el intelecto es separable (y la dificultad sobre las relaciones entre el intelecto activo y pasivo en *De Anima,* III, 4, hacen casi imposible ver si trataba de hacerlo). Pero sus seguidores han supuesto que era éste el argumento que le había llevado a escribir *De Anima,* 408b, 19-20 y 413b, 25 y ss., y a mí no se me ocurre nada mejor. Véase Mortimer Adler, *The Difference of Man and the Difference It Makes* (Nueva York, 1967), pág. 220, que cita 429a, 18-b23 como presagio del argumento tomista clásico de la explicación hilemórfica de la abstracción. Para una explicación deweyana de la «separabilidad» en cuanto «avena loca platónica», cfr. J. H. Randall, *Aristotle* (Nueva York, 1960), y el tratamiento hecho por Werner Jaeger a que se refiere Randall. Véase también Marjorie Grene, *A Portrait of Aristotle* (Londres, 1963), págs. 243 y ss. Comparto el desconcierto de Grene en este punto.

[9] Véase un punto de vista contrario interesante en T. H. Green, «The Philosophy of Aristotle» (en *Collected Works* [Londres, 1995], III, págs. 52-91). Green considera *De Anima,* III, 5 como un anticipo de Platón y la *Analítica Posterior* en orden al descubrimiento del holismo y del universal concreto. También piropea, incidentalmente, a Aristóteles (pág. 81) por apreciar la diferencia entre «sensación y conciencia inteligente de la sensación» que Locke no llegó a ver, por desgracia. Más adelante, y siguiendo a Kenny, señalo que el

más, no tiene sentido tratar de echar la culpa a Aristóteles o a sus intérpretes. La metáfora del conocimiento de las verdades generales por medio de la interiorización de los universales, de la misma manera que el ojo corporal conoce los particulares interiorizando sus colores y formas individuales, tenía, desde el mismo momento en que se presentó, fuerza suficiente para convertirse en el sustituto, entre los intelectuales, de la creencia de los campesinos en la vida entre las sombras. En formas diversas, que cubren la gama que va desde las nociones neoplatónicas del conocimiento en cuanto conexión directa con (emanación de, reflejo de) la divinidad, por una parte, hasta las versiones de la abstracción inspiradas en el hilemorfismo más prosaico del neo-aristotelismo, por la otra, la respuesta de los filósofos occidentales a la pregunta «¿Por qué es singular el hombre?» ha sido durante unos dos mil años: el alma en cuanto inmaterial-por-ser-capaz-de-contemplar-los-universales.

La tensión así establecida entre los dos lados de nuestro ser encontró una expresión convencional en pasajes como el discurso del «mono y la esencia» de Isabela:

Pero el hombre, el hombre orgulloso,
investido de una corta y débil autoridad,
conociendo menos aquello de que se cree más seguro,
es decir, su esencia de vidrio, parecido a un mono colérico
representa tan fantásticas comedias a la cara del Cielo,
que haría llorar a los ángeles, o,
si tuvieran el temperamento de nuestra naturaleza,
reír como mortales [10].

---

error de Locke fue consecuencia de la transformación cartesiana de la noción de mente.

[10] William Shakespeare, *Measure for Measure,* II, iii, 11, 117-123. Véase J. V. Cunnigham, «'Essence' and *The Phoenix and the Turtle*», *English Literary History,* 19 (1952), pág. 277, donde se afirma que esta «esencia de vidrio» es el «alma intelectual, que es de vidrio, pues refleja a Dios», pero Cunningham es convincente y le siguen los editores de *Arden Shakespeare* (a quienes debo la referencia a Cunnigham). En este punto, Shakespeare parece ser original, y no limitarse a usar una figura convencional. Parece que no hay ninguna alusión al pasaje del «*speculum obscurum*» de San Pablo ni a ninguna idea clásica. Para la historia de las analogías entre el alma y un espejo, véase Herbert Grabes, *Speculum, Mirror und Looking-Glass* (Tubinga, 1973), páginas 92 y ss. («Geistig-Seelisches als Spiegel»). La expresión *man's glassy essence [=esencia de vidrio del hombre]* fue mencionada por primera vez en filosofía por C. S. Peirce en un ensayo de 1892 con ese título donde se estudiaba la «teoría molecular del protoplasma», que, sorprendentemente, Peirce consideraba importante para confirmar la opinión de que «una persona no es nada más que un símbolo que implica una idea general» y para demostrar la

Nuestra Esencia de Vidrio —el «alma intelectual» de los escolásticos— es también la «mente del hombre» de Bacon que «muy diferente de la naturaleza de un vidrio claro y uniforme, en el que se reflejan los rayos de las cosas según su verdadera incidencia... es más bien como un espejo encantado, lleno de superstición e impostura, como no sea conjurado y reducido[11]. Estas ideas de comienzos del siglo XVII expresan una división dentro de nosotros mismos que se dejó sentir mucho antes de la Nueva Ciencia, de la división cartesiana entre sustancia pensante y extensa, del velo de las ideas y de la «filosofía moderna». Nuestra Esencia de Vidrio no era una doctrina filosófica, sino una imagen que los hombres de letras veían presupuesta en todo lo que leían. Es de vidrio —parecida a un espejo— por dos razones. En primer lugar, adopta nuevas formas sin cambiar —pero formas intelectuales, no formas sensibles como en el caso de los espejos materiales. En segundo lugar, los espejos están hechos de una sustancia que es más pura, de grano más fino, más sutil y más delicada que la mayoría[12]. A diferencia del bazo, que, junto con otros órganos igualmente toscos y visibles, explicaba la mayor parte de nuestra conducta, nuestra Esencia de Vidrio es algo que tenemos en común con los ángeles, aun cuando ellos lloren por nuestra ignorancia de su naturaleza. El mundo sobrenatural, para los intelectuales del siglo XVI, reproducía el modelo del mundo de las Ideas de Platón, lo mismo que nuestro contacto con él estaba inspirado en el modelo de su metáfora de la visión.

Hoy día son pocos los que creen en las ideas platónicas, y no hay muchos que establezcan una distinción entre el alma sensible y espiritual. Pero la imagen de nuestra Esencia de Vidrio sigue vigente, lo mismo que el lamento de Isabela por no poder compren-

---

existencia de «mentes de grupo» (cfr. *Collected Works,* ed. Charles Hartshorne y Paul Weiss [Cambridge, Mass., 1935], 6.270-171).

[11] Francis Bacon, *Advancement of Learning,* libro II en *Works,* ed. James Spedding y Robert Ellis (Boston, 1861), VI, 276.

[12] Cfr. *ibid.,* pág. 242, donde se trata de la afirmación de Bacon de que «el alma es la más simple de las sustancias». Cita en su favor un pasaje lucreciano tomado de Virgilio: *purumque reliquit/aethereum sensum atque auraï simplicis ignem* (Eneida, VI, 747). La idea de que el alma tiene que estar hecha de algún material especial de grano fino para ser capaz de conocimiento se remonta a Anaxágoras. La antigüedad vaciló entre el νοῦς en cuanto totalmente incorpóreo y en cuanto hecho de alguna materia especial, muy pura. Estas vacilaciones eran inevitables, dada la inimaginabilidad de lo no-espacio-temporal y la idea de que la razón debía parecerse a las verdades o formas no-espacio-temporales que comprende.

derla. A un sentimiento de fracaso moral se une una sensación de agravio porque la filosofía —la disciplina más interesada por «lo superior»— no nos ha hecho más conscientes de nuestra propia naturaleza. Se sigue teniendo la sensación de que esa naturaleza deja sentir más claramente su carácter distintivo en una forma determinada de conocimiento —el conocimiento de las cosas más altas y puras: las matemáticas, la misma filosofía, la física teórica, todo lo que contempla los universales. Sugerir que no existen los universales —que son *flatus vocis*— es poner en peligro nuestro carácter único. Decir que la mente es el cerebro es dar a entender que segregamos teoremas y sinfonías de la misma manera que nuestro bazo segrega malos humores. Los filósofos profesionales evitan estas «imágenes toscas» porque tienen otras imágenes —consideradas menos toscas— que fueron elaboradas a finales del siglo XVII. Pero sigue en pie la sensación de que la naturaleza de la razón es un «problema permanente» y que todo el que dude de nuestra singularidad debería estudiar matemáticas. El θυμός que estimulaba a los héroes homéricos, el ποεῦμ de San Pablo, y el entendimiento activo de Tomás de Aquino, son ideas muy diferentes. Pero, en este contexto, podemos fundirlas, como hace Isabela, en la expresión *Esencia de Vidrio*. Todas ellas son cosas que no tienen los cadáveres y que son distintivamente humanas. Los poderes manifestados por Aquiles no eran los de Teeteto o de los Apóstoles o de Santo Tomás, pero la «esencia intelectual» de los escolásticos heredaba las nociones dualistas que tomaron fuerza entre Homero y Anaxágoras, recibieron forma canónica en Platón, fueron moderadas por Aristóteles, y luego se vieron entremezcladas (en San Pablo) con un culto religioso nuevo y decididamente ultraterreno[13]. En las imágenes del «espejo» de los humanistas del Re-

[13] El vago dualismo de sentido común del cuerpo y el alma, de que pudo servirse Descartes, era producto del vocabulario de las traducciones vernáculas de la Biblia, al menos tanto como de cualquier otra causa. Por eso, para ver qué reciente y local es la distinción cartesiana, conviene tener en cuenta que los autores de la Biblia no estaban pensando en nada que se pareciera al contraste cartesiano entre «conciencia» y «materia insensible». Sobre las concepciones judías y su influencia en San Pablo, véase Onians, *Origins of European Thought*, págs. 480 y ss. Sobre el propio San Pablo es conveniente hacer notar que, a diferencia de los escritores modernos sobre filosofía de la mente, no identifica el cuerpo (σῶμα) con lo que se entierra después de la muerte. Esto último es σάρξ (carne) mientras que, según J. A. T. Robinson, «σῶμα es el equivalente más próximo a nuestra palabra 'personalidad'» *(The Body: A Study in Pauline Theology* [Londres, 1952], pág. 28; contrasta con Keith Campbell, *Body and Mind* [Nueva York, 1970], pág. 2: «Con tal de que sepas quién eres *tú*, es fácil decir qué es tu cuerpo: es lo que se entierra en el cementerio

nacimiento, se redujeron las diferencias entre Homero y Agustín, Plotino y Tomás, para producir un dualismo vago pero enfático —mono y esencia— que todos suponían formaba parte de la formación de los filósofos, aun cuando eran pocos los que llegaban a imaginarse lo que podría esperarse que dijeran al respecto. La actual filosofía de la mente suele incorporar a este vago conglomerado —la Esencia de Vidrio del hombre— las nociones post-cartesianas de «consciencia» o «conciencia». En la siguiente sección trataré de demostrar lo diferentes que son.

5. Capacidad de existir separadamente del cuerpo

El único punto de la sección anterior en que se ha utilizado un argumento ha sido la mención de la inferencia tomística (y quizá aristotélica) del carácter «separable», inmaterial del νοῦς a partir de la concepción hilemórfica del conocimiento —concepción según la cual el conocimiento no es la posesión de *representaciones* exactas de un objeto sino más bien que el sujeto se haga *idéntico* al objeto. Para ver la diferencia entre este argumento y los distintos argumentos cartesianos y contemporáneos en favor del dualismo, hay que ver hasta qué punto se diferencian estas dos epistemologías. Ambas utilizan las imágenes del Espejo de la Naturaleza. Pero en la concepción de Aristóteles el entendimiento no es un espejo examinado por un ojo interior. Es al mismo tiempo espejo y ojo en una misma cosa. La imagen de la retina es *ella misma* el modelo para el «entendimiento que se convierte en todas las cosas», mientras que en el modelo cartesiano, el entendimiento *examina* entidades que tienen como modelo a las imágenes de la retina. Las formas sustanciales de la «raneidad» y de la «estrelleidad» penetran hasta el entendimiento aristotélico, y están allí de la misma manera que lo están en las ranas y en las estrellas —*no* de la manera en que las ranas y estrellas aparecen reflejadas en los espejos. En la concepción cartesiana —que llegó a ser la base de la epistemología «moderna»— lo que hay en la «mente» son *representaciones*. El Ojo Interior

---

después de tu funeral»). Como dice Robinson (pág. 31n.) no es que σῶμα σάρξ sean *partes* distintas del hombre sino más bien «todo el hombre considerado de distinta manera». La idea del hombre dividido en *partes* no se da con naturalidad entre los no filósofos incluso después de Platón; véase van Peursen, *Body, Soul, Spirit*, cap. 6. Hay ejemplos de formas no cartesianas de utilizar σῶμα, σὰρξ, ψυχή y πνεῦμα por parte de Pablo en Corintios, I, 15:35-54.

examina estas representaciones con la esperanza de encontrar alguna señal que sea testimonio de su fidelidad.

Mientras que, en el mundo antiguo, el escepticismo había sido una cuestión de actitudes morales, un estilo de vida, una reacción a las pretensiones de las modas intelectuales del momento [14], el escepticismo al estilo de las *Primeras meditaciones* de Descartes era una pregunta perfectamente definida, precisa, «profesional»: ¿Cómo sabemos que lo que es mental representa algo que no es mental? ¿Cómo sabemos si lo que ve el Ojo de la Mente es un espejo (aun cuando sea un espejo deformado —un espejo encantado—) o un velo? La idea del conocimiento en cuanto representación interna nos es tan natural que el modelo de Aristóteles nos parece simplemente pintoresco, y el escepticismo cartesiano (en oposición al escepticismo «práctico» de Pirrón) nos parece tan esencial a lo que significa «pensar filosóficamente» que nos sorprende que Platón y Aristóteles lo consideraran directamente. Pero si vemos que los dos modelos —el hilemórfico y el representativo— son igualmente opcionales, quizá podamos ver las inferencias para el dualismo cuerpo-mente que se derivan de cada uno como meramente opcionales.

En un artículo titulado «Why Isn't the Mind-Body Problem Ancient?» [15] [¿Por qué no es antiguo el problema mente-cuerpo?],

[14] Véase Philip P. Hallie sobre el escepticismo griego en cuanto «filosofía eudomonística de la sabiduría práctica» cuya «duda, más que un instrumento para apartar el velo de la experiencia sensorial, es un medio de eliminar las excrecencias que ensucian la vida del hombre y la llevan a conflictos interminables y amargos con los demás hombres» (*Scepticism, Man and God* [Middletown, Conn., 1964], pág. 7). No es claro qué papel desempeñó la idea del velo de las ideas en el escepticismo antiguo, pero parece que fue incidental, y no central, en el sentido en que lo convirtieron en central Locke-Berkeley-Kant. Charlotte Stough (Greek Scepticism [Berkeley, 1969], pág. 24) dice de Pirrón que considera τὸ φαινόμενον como «una cortina entre el sujeto y el objeto, que oculta de su vista el mundo real». Sin embargo, no está claro que το φαινόμενον se parezca mucho a una idea lockeana, que está incorregiblemente ante la mente, y es puramente mental sencillamente porque es conocida incorregiblemente. Lo más próximo del pensamiento antiguo a la idea de una clase de entidades mentales incorregiblemente conocibles es la doctrina estoica de κάϊαληπῖικὴ φαντασία (cfr. Stough, págs. 38-40), pero ésta se define como una representación que se corresponde exactamente con su objeto y, *por tanto,* obliga a asentir, lo cual dista mucho de ser la idea de Locke. Véase también Josiah B. Gould, «Being, the World and Appearance in Early Stoicism and Some Other Greek Philosophers», *Review of Metaphysics,* 27 (1974), 261-288, especialmente págs. 277 y ss.

[15] Wallace Matson, «Why Isn't the Mind-Body Problem Ancient?» en *Mind, Matter and Method: Essays in Philosophy and Science in Honor of Herbert Feigl,* ed. Paul Feyerabend y Grover Maxwell (Minneapolis, 1966), páginas 92-102.

Wallace Matson ha señalado el punto principal que distingue la forma en que consideraron la separación de la mente y el cuerpo los griegos y los hombres del siglo XVII:

> Los griegos no carecían de un concepto de mente, incluso de una mente separable del cuerpo. Pero desde Homero hasta Aristóteles, la línea divisoria entre mente y cuerpo, si es que llegaba a trazarse, se trazaba de forma que incluyera los procesos de la percepción sensorial en la parte del cuerpo. Esa es una de las razones por las que los griegos no tuvieron ningún problema mente-cuerpo. Otra razón es que resulta difícil, casi imposible, traducir al griego una oración como «¿Cuál es la relación de la sensación con la mente (o el alma)?». La dificultad estriba en encontrar un equivalente griego de «sensación» en el sentido en que la utilizan los filósofos... El término «sensación» se introdujo en la filosofía precisamente para que fuera posible hablar de un estado consciente sin comprometerse sobre la naturaleza ni siquiera sobre la existencia de los estímulos externos [16].

Se pueden resumir las dos precisiones de Matson diciendo que en griego no es posible separar los «estados conscientes» o «estados de conciencia» —hechos de una vida interior— de los hechos del «mundo externo». Por otro lado, Descartes utilizó el término «pensamiento» para referirse al dudar, comprender, afirmar, negar, querer, rechazar, imaginar y sentir, y decía que, incluso si sueño que veo la luz, «propiamente hablando lo que hay en mí se llama sensación, y se utiliza en este sentido preciso de que no es otra cosa que pensamiento» [17]. Una vez que Descartes hubo introducido fir-

---

[16] Matson, «Mind-Body Problem», pág. 101. Luego pasa a señalar que ni σισθησις ni αἴσθημα pueden figurar como equivalentes de «sensación». φάντασμα es una posibilidad tentadora, pero ni siquiera el uso que Aristóteles hace del término se puede traducir por «imagen mental» sin provocar sospechas, y nadie diría que un dolor es un φαντασμα. Sobre la dificultad de interpretar la idea de *fantasma* de Aristóteles, véase Anthony Kenny, «Intellect and Imagination in St. Thomas» en *Aquinas: A Collection of Critical Essays*, ed. Kenny (Garden City, Nueva York, 1969), págs. 293-294. La afirmación de Matson en la frase final de esta cita la respalda Thomas Reid en su interpretación del término «sensación». Véanse sus *Essays on the Intellectual Powers of Man* (Cambridge, Mass., 1969), pág. 249 (y el ensayo II, *passim)*.

[17] En la Meditación II, Descartes comienza definiendo una «cosa que piensa» como «una mente o alma, o un entendimiento, o una razón *(res cogitans, id est, mens, sive animus, sive intellectus, sive ratio)* y añade en seguida: «¿Qué es una cosa que piensa? Es una cosa que duda, entiende, afirma, niega, quiere, rechaza, *y también imagina y siente* (el subrayado es mío; «*Nempe dubi-*

memente esta forma de hablar. Locke pudo utilizar el término «idea» en un sentido que no tiene ningún equivalente griego, con el significado de «todo lo es el objeto de comprensión cuando piensa un hombre» o «todo objeto inmediato de la mente al pensar» [18]. Como dice Kenny, el uso moderno de la palabra *idea* procede de Descartes a través de Locke, «y Descartes le dio conscientemente un nuevo

---

tans, intelligens, affirmans, negans, volens, nolens, imaginans quoque, et sentiens»). Luego continúa con el segundo fragmento citado más arriba *(hoc est proprie quod in me sentire appelatur; atque hoc praecise sic sumptum nihil aliud est quam cogitare)*. Los tres textos están en las págs. 184-186 de *Oeuvres Philosophiques,* ed. Alquié (París, 1967), vol. II (págs. 152-153 del vol. I de la traducción de Haldane y Ross). Véase también *Principles,* I, sección 9: «Con la palabra *pensamiento* me refiero a todo aquello de que somos conscientes que opera en nosotros *(tout ce qui se fait en nous de telle sorte que nous l'apercevons immédiatement par nous-mêmes)*. Y ésa es la razón por la que en esto no sólo el entender, querer e imaginar sino también el sentir *(sentir)* son la misma cosa que el pensamiento» *(Oeuvres Philosophiques,* III, pág. 95; Haldane y Ross, I, pág. 222). Sobre la traducción de *res cogitans* como «conciencia» véase Robert McRae, «'Descartes' Definition of Thought» en *Cartesian Studies,* dirigido por R. J. Butler (Oxford, 1972), páginas 55-70.

[18] La primera cita está tomada de *Essay* (I, i, 8) y la segunda de la «Segunda carta al obispo de Worcester». La inmediatez en cuanto señal de lo mental (siendo la incorregibilidad el criterio de la inmediatez) llegó a adquirir en la filosofía el carácter de presuposición indiscutida gracias a textos como éstos. Y, con mucha frecuencia, en filosofía el uso neologista se convirtió en señal de una comprensión de los temas y problemas «distintivamente filosóficos». Así, vemos cómo Hume dice. «... todas las conclusiones que el vulgo coloca en este apartado son directamente contrarias a las confirmadas por la filosofía. La filosofía nos informa de que todo lo que aparece ante la mente no es nada más que una percepción y es interrumpido y depende de la mente; mientras que el vulgo confunde las percepciones y los objetos, y atribuye una existencia continuada y distinta a las mismas cosas que sentimos y vemos» *(Treatise,* I, iv, ii). Jonathan Bennett comenta que «el pensamiento de Locke está dominado por su intento de utilizar unívocamente 'idea' como término clave de sus explicaciones de la percepción y del significado —o, dicho en pocas palabras, su uso de 'idea' para referirse tanto a los conceptos como a los datos sensoriales» y dice que en esto «se personifica su error fundamental, compartido con Berkeley y Hume y otros autores dentro de la tradición empirista, de asimilar demasiado íntimamente lo sensorial y lo intelectual» (Benhett, *Locke, Berkeley, Hume: Central Themes* [Oxford, 1971], pág. 25). Este error, sin embargo, se remonta hasta Descartes y se manifestó también en la tradición racionalista. Forma parte de lo que Wilfrid Sellars llama el «marco de lo dado», común a ambas tradiciones, y ha sido siempre el blanco de los que han recibido la influencia de Hegel. Cfr. Sellars, *Science, Perception and Reality* (Londres y Nueva York, 1963), págs. 127 y 155-156. Las quejas de Sellars y Bennett encuentran un precedente en H. L. Prichard, y en fecha todavía más temprana, en T. H. Green, de quien hablo brevemente en el capítulo tercero, sección 2.

sentido... le dio una nueva orientación al usarlo sistemáticamente para los contenidos de una mente humana»[19]. Y lo que es más importante, no había habido ningún término, ni siquiera en el campo de la filosofía, dentro de las tradiciones griega y medieval que fuera coextensivo con el uso de «idea» por Descartes-Locke. Tampoco se había dado la concepción de la mente humana en cuanto espacio interior en el que los dolores y las ideas claras y distintas pasaban revista ante un único Ojo Interior. Existían, ciertamente, las nociones de pensamiento tácito, de resoluciones *in foro interno,* y semejantes[20]. La novedad estuvo en la idea de un solo espacio interior en el que eran objeto de cuasi-observación las sensaciones corporales y perceptivas («ideas confusas del sentido y la imaginación», como diría Descartes), las verdades matemáticas, las reglas morales, la idea de Dios, los talantes depresivos, y todo el resto

[19] Anthony Kenny, «Descartes on Ideas» en *Descartes: A Collection of Critical Essays,* ed. Willis Doney (Garden City, Nueva York, 1967), pág. 226. Véanse las definiciones de Descartes de «pensée» como «tout ce qui est tellement en nous, que nous en sommes immédiatement connaissants» y de «idée» en cuanto «cette forme de chacune de nos pensées, par la perception inmédiate de laquelle nous avons connaissance de ses mêmes pensées» *(Replies to Second Objections,* edición Alquié, II, 586). No obstante, John Yolton se pone de parte de Kenny (y con Alquié y otros comentaristas que siguen la línea tradicional —que yo estoy adoptando aquí— de que la doctrina de Descartes de la percepción representativa era un corte radical, y quizá desastroso, con la tradición escolástica del realismo directo). En sus «Ideas and Knowledge in Seventeenth-Century Philosophy» *(Journal of the History of Philosophy,* 13 [1975], 145-165) cita la descripción que hace Descartes de «idée» como «una manière ou façon de penser» en la *Tercera Meditación* (Alquié, II, 439) como prueba de que Descartes tenía una teoría de las ideas que era compatible con el realismo directo escolástico. En ésta y en otras obras Yolton ha dado a entender que quizá sea demasiado simplista la versión más frecuente (compartida, por ejemplo, por Étienne Gilson y J. H. Randall) sobre la aparición del escepticismo epistemológico a partir de una teoría de la percepción representativa que sería obra de Descartes y Locke. En *Epistemological Direct Realism* (Alburquerque, Nueva York, 1974), págs. 96-97, Brian O'Neil sigue una línea parecida: «La larga lucha mantenida por Descartes fue un esfuerzo por conservar y relacionar la teoría del *esse objectivum* y la doctrina de las naturalezas simples.» O'Neill está de acuerdo con Jean Wahl en que Descartes «ait exprimé les deux conceptions fondamentales et antinomiques du réalisme», una basada en algo parecido al tomismo y la otra basada en algo parecido a un velo de ideas, algunas de las cuales garantizan su propia precisión en cuanto representaciones. Si las interpretaciones y revisiones que Yolton hace de Locke y Descartes están en lo cierto, habrá que seguir buscando en la historia para dar con la aparición de lo que ahora se considera como la problemática epistemológica creada por Descartes. Sin embargo, en estas páginas estoy siguiendo la versión, más conocida, de Kenny.

[20] Véase, por ejemplo, Platón, *Sofista,* 263E.

de lo que llamamos «mental». Este escenario interior, con su observador interno, había aparecido insinuado en varios momentos del pensamiento antiguo y medieval pero nunca se había tomado en serio el tiempo suficiente como para servir de base a una problemática [21]. Pero el siglo XVII se tomó con la seriedad necesaria para permitir que se planteara el problema del velo de las ideas, el pro-

---

[21] Adler *(Difference of Man,* págs. 217-218) está de acuerdo con Matson en que «dentro del marco de la psicología y metafísica aristotélica no puede haber ningún problema mente-cuerpo» pero afirma que «Platón, por ejemplo, habría entendido a Descartes mucho mejor que Aristóteles, especialmente la separación cartesiana de la mente y el cuerpo en sustancias existencialmente distintas y la concepción cartesiana de la independencia de la mente y el cuerpo». De todas formas, me atrevería a poner en duda que haya una diferencia real entre Platón y Aristóteles en este punto; la indicación de Matson sirve con la misma perfección para los dos. Por otra parte, he de admitir que hay algo de cierto en la opinión de Gilson de que Descartes recupera precisamente aquellos elementos de la tradición agustiniana que Santo Tomás había criticado, sirviéndose de Aristóteles. Por eso la opción de Coleridge entre Platón y Aristóteles parece pertinente. Además, hay párrafos de Agustín que están muy cerca de algunos pasajes de Descartes que se citan frecuentemente para demostrar que la originalidad de su noción del «pensar» se refería al sentido y al entendimiento. Gareth Matthews (en «Consciousness and Life», *Philosophy,* 52 [1977], páginas 13-26) cita un ejemplo sorprendente tomado de *Contra Académicos,* libro 3, sección 11, capítulo 26, y comenta:

> La imagen de los seres humanos en cuanto dotados... tanto de un «interior» como de un «exterior» es tan corriente, tan de sentido común (según nuestra forma de ver las cosas), que nos cuesta trabajo darnos cuenta de que es sorprendentemente moderna. Pero para apreciar su modernidad sólo hay que buscar formulaciones de la misma antes de Descartes. Se encuentran algunas anticipaciones interesantes en Agustín, pero es poco lo que aparece antes y poco lo que encontraremos entre la época de Agustín y la de Descartes (página 25).

La precisión de Matthews (como la de Matson) me parece un correctivo útil de la afirmación de que Ryle estaba atacando una intuición humana básica cuando atacaba la idea del fantasma de la máquina, y no simplemente algo peculiar de Descartes. Esto es lo que afirma, por ejemplo, Stuart Hampshire («Critical Study» de *The Concept of Mind, Mind,* 59 [1950], 237-255, especialmente la sección 2). Por otra parte, la crítica de Hampshire se confirma por la indicación (que me hizo Michael Frede en una conversación) de que la aparente novedad de las doctrinas cartesianas de la percepción representativa y del «espacio interior» de los seres humanos pierde fuerza si se lee la filosofía helenística y se aprecia el papel de los estoicos en el pensamiento renacentista. Si Frede está en lo cierto, y especialmente si Yolton también lo está en las cuestiones mencionadas más arriba en la nota 19, hay muchas más continuidades en la historia del examen filosófico de estos temas de lo que se refleja en la historia tal como la estoy contando (inspirada en la tradición historiográfica de Gilson-Randall).

blema que dio a la epistemología un papel central en la filosofía.

Una vez que Descartes hubo inventado ese «sentido preciso» del «sentir» en que «no era otra cosa que pensar», comenzamos a perder contacto con la distinción aristotélica entre la razón-en-cuanto-comprensión-de-los-universales y el cuerpo vivo que se encarga de la sensación y el movimiento. Hacía falta una *nueva* distinción mente-cuerpo —la que llamamos distinción «entre la conciencia y lo que no es conciencia». No se trataba de una distinción entre facultades humanas sino de una distinción entre dos series de hechos, de manera que muchos hechos de una serie tenían muchas características iguales a las de muchos hechos del otro, pero a pesar de ello se diferenciaban *todo caelo* porque uno era un hecho en una sustancia extensa y el otro en una sustancia no extensa. Se trataba más de una distinción entre dos mundos que de una distinción entre dos lados o partes de un ser humano. El «mundo Ideal» de filósofos como Royce hereda el prestigio y el misterio de la Esencia de Vidrio del Renacimiento, pero es autosuficiente en una forma en que no podría serlo nunca una parte de un hombre [22]. Demostrar que la mente podía imaginarse aparte del cuerpo era, por tanto, un proyecto totalmente diferente del que se encuentra en la tradición procedente de Aristóteles. En Aristóteles, la facultad que recibía los universales sin darles forma material era «separable» y no se podía saber fácilmente (sin ayuda de intereses extrafilosóficos, como el Cristianismo) si había que considerarla como un poder especial que tenía el cuerpo, una sustancia separada unida a cada

---

[22] Véase la descripción de Royce del subjetivismo cartesiano como «redescubrimiento de la vida interior» («The Rediscovery of the Inner Life», capítulo III de su *The Spirit of Modern Philosophy* [Nueva York, 1892]) y como apertura a una comprensión de que el mundo real debe ser «mental» (conclusión que expone triunfalmente en el capítulo XI). Véase también el capítulo primero de Lovejoy, *The Revolt Against Dualism* (La Salle, Ill., 1930), en que Lovejoy insiste, contra los que quieren derrocar a Descartes, en que el velo de ideas es un problema que se presenta a todos los que conservan

> la fe primaria y más universal del hombre, su realismo inexpugnable, su doble creencia de que, por un lado está en medio de realidades que no son ni él mismo ni meras sombras obsequiosas de sí mismo, un mundo que transciende los estrechos confines de su ser pasajero; y, por el otro lado, que puede de alguna manera llegar más allá de esos confines y hacer entrar estas existencias externas en el círculo de su propia vida, sin anular por ello su transcendencia (página 14).

A Aristóteles, Tomás de Aquino, Dewey o Austin este «realismo» les parecería tan artificial y forzado como el idealismo de Royce.

cuerpo humano maduro, o quizá una sustancia única que era compartida de alguna manera por todos los hombres y ángeles que pudieran existir. Aristóteles vacilaba entre la primera y la segunda opción, teniendo la segunda los atractivos que ofrece generalmente la posibilidad de sobrevivir a la muerte. La filosofía medieval vaciló entre la segunda y la tercera. Pero en todas estas disputas, la contorversia no se refería a la supervivencia de la «conciencia» sino a la indestructibilidad de la *razón* [23]. Desde el momento en que la

[23] Así, Agustín inicia su tratado «Sobre la inmortalidad del alma», con lo que considera el argumento más simple y decisivo de todos: el alma es inmortal porque es el sujeto (es decir, la sede) de la ciencia, que es eterna. En el capítulo II, por ejemplo, dice: «El cuerpo humano es mutable y la razón es inmutable. Todo lo que no existe siempre de la misma forma es mutable, pero que dos y dos son cuatro existe siempre de la misma forma... Esta forma de razonamiento es, por tanto, inmutable. Por consiguiente, la razón es inmutable» (*Concerning the Teacher* y *On the Immortality of the Soul,* traducción de George Leckie [Nueva York, 1938], pág. 61). Entre el *Fedón,* de Platón, y el siglo XVII, el argumento *filosófico* clásico para demostrar la inmortalidad se había centrado siempre en torno a nuestra capacidad de hacer lo que no pueden hacer las bestias —conocer verdades inmutables, y no sólo hechos concretos. Incluso Descartes, que abrió las puertas a una concepción totalmente nueva de la diferencia entre mente y cuerpo, se inclinaba a volver a la posición clásica y decía que el cuerpo era responsable de todas las acciones que compartimos con las bestias —por ejemplo, huir del peligro. Así en la *Respuesta Cuarta* dice que el acto reflejo se produce «sin ayuda de la mente», y que esto no debe producir más extrañeza que el que «la luz reflejada del cuerpo de un lobo en los ojos de una oveja sea capaz de excitar un movimiento de huida» (*Oeuvres Philosophiques,* ed. Alquié, II, 671; Haldane y Ross, II, 104). Pero parecería que su tratamiento del sentimiento como una forma de pensamiento le obligaría a hacer la afirmación paradójica (que ni Aristóteles ni Agustín tendrían motivos para hacer) de que el sentimiento de terror que acompaña a nuestra huida no tiene paralelo en la oveja. Véase el argumento en una carta a Henry More (5 de febrero de 1649), pág. 885 del vol. III, de la edición Alquié. En pasajes como éste, llega a un punto decisivo la ambigüedad de «pensamiento» entre «razonamiento» y «conciencia», Descartes necesita el primer sentido para evitar la paradoja y mantener contacto con la tradición y el segundo sentido para establecer un dualismo de la sustancia extensa y no extensa. Puede verse una revisión de los textos cartesianos relevantes y una buena exposición del impacto de la concepción de Descartes sobre la filosofía posterior en Norman Malcolm, «Thoughtless Brutes», *Proceedings and Addresses of the American Philosophical Association,* 46 (1973), 5-20. Sin embargo, Malcolm piensa que mi pregunta de por qué agrupó Descartés los dolores y pensamientos para incluirlos dentro de la misma sustancia se responde diciendo que lo que hay de común en imaginar, querer, sentir, percibir, etc., es que en todos ellos «hay un objeto de la conciencia» (Malcolm, «Descartes's Proof That His Essence Is Thinking», en *Descartes: A Collection of Critical Essays,* ed. Willis Doney, pág. 317n). En otras palabras, Malcolm piensa que la intencionalidad es suficiente para unificar todas las cosas que Descartes desea unir bajo *cogitatio* y *pensée.*

mente ya no es sinónimo de la razón, lo que sirva de marca distintiva de la mente ha de ser algo distinto de nuestra comprensión de las verdades universales.

Si buscamos en Descartes un factor común entre los dolores, sueños, imágenes de la memoria, y percepciones verídicas y alucinatorias, por una parte, y, por la otra, los conceptos de (y juicios acerca de) Dios, el número y los constitutivos últimos de la materia, no buscamos ninguna doctrina explícita. Descartes nos dice que comprendemos claramente la distinción entre lo extenso y lo no extenso, y es cierto (en el mismo sentido superficial en que podríamos hablar de que comprendemos claramente la distinción entre lo finito y lo infinito), pero esto no sirve de ayuda para los casos límite (captación sensorial de los particulares) que constituyen, casualmente, el núcleo de la cuestión. Es precisamente la condición de las «ideas confusas del sentido y la imaginación» lo que establece la diferencia entre la mente-en-cuanto-razón y la mente-en-cuanto-conciencia.

La respuesta que me gustaría dar a la pregunta «¿Qué factor común encontró Descartes?» es «la indubitabilidad», es decir, el hecho de que los dolores, igual que los pensamientos y la mayoría de las creencias, son de tal condición que el sujeto no puede dudar que los tiene, mientras que es posible dudar de todo lo que es físico. Si presentamos esta respuesta, podemos considerar lo que Royce llamó «redescubrimiento cartesiano de la vida interior» como el descubrimiento de la verdadera esencia de la conciencia —que no hay distinción entre apariencia y realidad, mientras que en todo lo demás la hay. No obstante, el problema de dar esta respuesta está en que nunca la propone explícitamente el propio Descartes. Por eso, lo mejor que puedo hacer para justificarla es decir que hace falta *algo* para explicar el reagrupamiento que hace Descartes de los distintos elementos que Aristóteles y Tomás de Aquino habían separado, que no hay ninguna otra cosa que pueda servirnos para ello, y que la indubitabilidad era tan querida para el autor de la

---

No creo que esto sirva en el caso de los dolores. De todas formas, une la intencionalidad auténtica de las representaciones lingüísticas y la pseudo-intencionalidad (como la llama Sellars) de las sensaciones que Tomás de Aquino mantenía separadas. Es precisamente esta fusión de Descartes lo que hay que explicar. O, dicho en otras palabras, lo que hay que explicar es precisamente el origen de la idea de «conciencia» en la expresión «objeto de conciencia». Sobre la distinción entre intencionalidad auténtica y pseudo-intencionalidad, véase Sellars, «Being and Being Known», en *Science, Perception and Reality.*

Primera Meditación que parece constituir un motivo natural de esta revolución conceptual.

Margaret Wilson ha señalado que podemos encontrar en Descartes un argumento en favor del dualismo mente-cuerpo (dualismo que sigue las líneas revisionistas que vengo describiendo) que es un sencillo «argumento a partir de la duda». Este argumento dice que lo que podemos dudar que existe no puede, por la ley de Leibniz, ser idéntico a lo que no podemos dudar que existe. Como dice Wilson, este argumento «es reconocido universalmente como falaz»[24]. Es falaz, cuando menos, porque la ley de Leibniz no se aplica a las propiedades intencionales. Pero, sigue diciendo Wilson, el argumento en favor del dualismo que aparece en la *Sexta Meditación no* es una versión de este argumento falso. Se centra en la idea de «una cosa completa» (que parece lo mismo que la idea de «sustancia» en el sentido del término en que Descartes admitirá únicamente tres sustancias —el pensamiento, la extensión y Dios). La premisa fundamental es «Puede entender de forma clara y distinta que algo puede ser una cosa completa si tiene χ (una propiedad psicológica) aun cuando carezca de ϕ (una propiedad física)» (pág. 14).

Creo que el análisis de Wilson es correcto, y que tiene también razón cuando dice que este argumento en su conjunto «no es mejor que la distinción entre percepción clara y distinta y 'mera' percepción», en relación con el cual la autora duda de que «las llamadas de los esencialistas recientes a la intuición» tengan una base más consistente (pág. 14). No obstante, mi opinión personal es que la cuestión más importante suscitada por su análisis es cómo Descartes consiguió convencerse a sí mismo de que algo que incluía tanto los dolores como el conocimiento matemático era «una cosa completa» y no dos cosas. A su vez, esto se reduce a la cuestión de cómo fue capaz de dar al término *penser* el sentido extenso de «conciencia» al mismo tiempo que lo consideraba como el nombre de una sustancia separada, de la misma manera en que νοῦς e *intellectus* se habían difundido como nombres de sustancias separadas. Mi opinión particular, tal como he señalado, es que las «intuiciones esencialistas» y «las percepciones claras y distintas» son siempre llamamientos a hábitos lingüísticos enraizados en el lenguaje por nuestros predecesores. Por eso, lo que hay que explicar es cómo pudo convencerse a Descartes de que su reagrupación era «intuitiva».

---

[24] Véase Margaret Wilson, «Descartes: the Epistemological Argument for Mind-Body Distinctness», *Nous*, 10 (1976), 7-8. Debo agradecer a Wilson sus observaciones, sumamente valiosas, sobre un borrador de este capítulo.

Admitiendo que el «argumento a partir de la duda» no tiene ningún valor, creo, sin embargo, que es uno de esos casos en que «se presentan razones sin valor para justificar lo que creemos por instinto» y que sirve de pista de los instintos que han sido los verdaderos responsables de la convicción. El presentimiento que se daba en este caso era, creo yo, que las verdades matemáticas conocidas indubitablemente (una vez desarrolladas las pruebas para que sean percibidas clara y distintamente con una especie de intensidad «fenoménica» y en forma no discursiva) y los indudables estados de conciencia momentáneos tenían algo en común —algo que hacía posible agruparlos en una sustancia única. Así, Descartes dice:

> Ahora bien, en lo que se refiere a las ideas, si las consideramos únicamente en sí mismas y no las relacionamos con nada que esté más allá de ellas, no pueden, propiamente hablando, ser falsas; pues tanto si imagino una cabra como si imagino una quimera, no es menos cierto que imagina la una y no la otra [25].

Hablando de los seres imaginarios creados por los pintores, dice que aun cuando

> su obra representa una cosa puramente ficticia y absolutamente falsa, es cierto, sin embargo, que los colores de que se compone son necesariamente reales. Y por la misma razón, aunque las cosas generales *(generalia)* como los ojos, la cabeza, las manos, etc., puedan ser imaginarias, es necesario que haya algunas cosas más simples y universales *(quaedam aduc magis simplicia et universalia)* de las cuales estén hechas —como están hechas con colores todas las imágenes de las cosas *(rerum imagines)*, sean verdaderas o falsas, que están en nuestro pensamiento *(quae in cogitatione nostra sunt)* [26].

En estos pasajes, me parece que Descartes está vislumbrando vagamente una semejanza entre las «naturalezas simples» que conocemos en la física matemática (que pueden ser las *quadam simplicia et universalia* en cuestión) y los colores mismos. Los colores, en su concepción metafísica oficial, galileana, son cualidades secundarias, que están a la espera de ser analizadas en otras más simples, pero parece que epistemológicamente tienen, como el dolor, el mismo tipo de inevitabilidad primitiva que las mismas naturalezas simples. No podía explicitar la analogía sin dar un paso hacia el empirismo

---

[25] *Meditation III,* edición Alquié, II, 193.
[26] *Meditation I,* edición Alquié, II, 179.

lockeano. Pero tampoco podía renunciar a ella sin reincidir en las antiguas distinciones aristotélicas entre las almas sensible e intelectual. Esto habría supuesto una vuelta a la metafísica anterior a Galileo, cosa que él quería evitar, por no hablar de una epistemología hilemórfica que difícilmente se podía reconciliar con la capacidad explicativa de la mecánica galileana[27]. Ante esta difícil situación, dejó, en mi opinión, que la mayor parte de la tarea de cambiar la idea de «mente» se hiciera a escondidas, no mediante un argumento explícito sino con maniobras verbales que barajaban las cartas de la baraja de forma ligeramente distinta en cada pasaje en que aparecía la distinción entre mente y cuerpo[28].

Si estoy en lo cierto al afirmar que la corazonada de Descartes que no llegó a demostrar, la que le permitió ver los dolores y pensamientos como modos de una sola sustancia, fue que la indubitabilidad era el carácter común que no compartían con nada físico, entonces podemos verle abrirse paso hacia una concepción en que la indubitabilidad no sea ya señal de eternidad, sino más bien de algo para lo que los griegos no tenían nombre: la conciencia. Mientras que los filósofos anteriores habían seguido más o menos a Platón al pensar que sólo se conocía con certeza lo eterno, Descartes

---

[27] Véase el comentario de A. G. A. Balz:

> Creo que a Descartes le habría gustado afirmar que el intelecto es la única cosa pensante, la sustancia del alma inmaterial e inextensa, que atribuye al cuerpo imaginación, percepción y sentimiento. Pero si el dolor, causado por un cuchillo no es propiedad del cuchillo en cuanto constelación de modos de la materia, no puede ser una propiedad de esa constelación de modos de materia que es el cuerpo humano. Por eso, el dolor y el resto de nuestra experiencia inmediata deben colocarse, por fuerza, del lado de la sustancia del alma («Concerning the Thomistic and Cartesian Dualisms: A Rejoinder to Professor Mourant», *Journal of Philosophy*, 54 [1957], 387).

[28] Ese juego de manos inconsciente, cuando lo practican hombres con una imaginación tan audaz como la de Descartes, constituye motivo de agradecimiento, más que de crítica. Ningún gran filósofo lo ha evitado, y ninguna revolución intelectual ha podido triunfar sin él. En terminología «kuhniana», no puede triunfar ninguna revolución que emplee un vocabulario conmensurable con el anterior, y por eso nadie puede triunfar empleando argumentos que utilicen inequívocamente términos compartidos con la sabiduría tradicional. Por eso, los malos argumentos para demostrar corazonadas brillantes deben preceder necesariamente a la normalización de un nuevo vocabulario que incorpore tal corazonada. Dado ese nuevo vocabulario, son posibles argumentos mejores, aunque las víctimos de la revolución los considerarán siempre una petición de principio.

estaba reemplazando con la «percepción clara y distinta» —es decir, la clase de conocimiento no confuso conseguido mediante un proceso de análisis— a la «indubitabilidad» en cuanto señal de las verdades eternas. *Con ello la indubitabilidad quedaba libre para servir como criterio de lo mental.* Aunque el pensamiento de que tengo dolor no figura como una percepción clara y distinta, no puede ponerse en duda con más razones que el pensamiento de que existo. Mientras que Platón y la tradición habían hecho coincidir las líneas entre confusión y claridad, dubitabilidad e indubitabilidad, y mente y cuerpo, Descartes se disponía a redistribuirlas. El resultado fue que de Descartes en adelante tenemos que distinguir entre la razón metafísica especial de nuestra certeza sobre nuestros estados internos («nada está más cerca de la mente que ella misma») y las diversas razones epistemológicas que sirven de base a nuestras certezas sobre cualquier otra cosa. Esta es la razón por la que, una vez que se trazó claramente esta distinción, y se disipó la confusión del propio Descartes entre la certeza de que existe algo y la certeza sobre su naturaleza, el empirismo comenzó a arrinconar al racionalismo. Nuestra certeza de que nuestro concepto de «doloroso» o «azul» significa algo real acaba con nuestra certeza de que tenemos una percepción clara y distinta de las naturalezas simples, como pueden ser «sustancia», «pensamiento» y «movimiento». Con el empirismo lockeano, la epistemología, con carácter de ciencia fundamental, pasó a convertirse en el paradigma de la filosofía [29].

El propio Descartes intentó continuamente mantenerse fiel a las distinciones platónicas y escolásticas clásicas, por una parte, al tiempo que las desmontaba, por la otra. Así vemos cómo, cuando se tiene que enfrentar con Hobbes [30], utiliza la glándula pineal para

[29] El descubrimiento de la posibilidad de un fundacionalismo empirista está vinculado a lo que Ian Hacking ha descrito en *The Emergence of Probability* (Cambridge, Cambridge University Press, 1975) como invención de la idea de «evidencia» en el sentido moderno —idea que era un prerrequisito de los proyectos fundacionalistas, y a fortiori del empirismo. Esta invención, así como el triunfo final del empirismo, iba unida a la distinción entre ciencias superiores e inferiores (cfr. Hacking, pág. 35, y T. S. Kuhn, *The Essential Tension* [Chicago, 1978], capítulo 3). Una explicación más completa de estos cambios cartesianos uniría en uno estos temas.

[30] Véase su afirmación de que Hobbes confunde las ideas propiamente dichas con «les images des choses matérielles depeinte en la fantasie corporelle», que sería la glándula pineal, en *Replies to Third Objections,* edición Alquié, II, 611. Margaret Wilson, en «Cartesian Dualism» (en *Descartes: Critical and Interpretive Essays,* edición Michael Hooker, Baltimore, 1978) apunta, basándose en estos pasajes, que deberíamos ser cautos al atribuirle la opinión de que «percibimos clara y distintamente nuestras sensaciones aparte

reintroducir la distinción entre el alma sensible y el alma intelectiva, y vuelve a utilizarla para recrear la clásica asociación paulina entre las pasiones y la carne, en *Las pasiones del alma.* Pero este doble juego fue ridiculizado por algunos, por ejemplo Spinoza, que veía claramente que una idea confusa, pero puramente mental, podía hacer todo lo que pudieran hacer los espíritus animales o «la fantasie corporelle» [31]. Esta segunda generación de cartesianos, que consideraban que el propio Descartes no había conseguido liberarse totalmente del fango escolástico, habían purificado y «normalizado» la doctrina cartesiana, y se llegó así a la versión plenamente desarrollada de la «idea 'idea'», que hacía posible que Berkeley pensara que la sustancia extensa era una hipótesis que no necesitaba para nada. Este pensamiento no se le habría ocurrido nunca a un obispo pre-cartesiano, en pugna con la carne más que con la confusión intelectual. Con esta «idea 'idea'» plenamente desarrollada llegó la posibilidad de la filosofía en cuanto disciplina que se centraba nada menos que en la epistemología, y no en Dios y la moralidad [32].

---

de todo acontecimiento o estado físico». Hay que tener cautela, ciertamente, pero esto no significa que sea posible hacer que el rechazo por Descartes de esta afirmación concuerde con los pasajes dualistas más «centrales» de las *Meditations.* (Este ensayo de Wilson hace también la afirmación interesante de que el mismo Descartes, a diferencia de Gassendi, Hobbes y Spinoza, no creía en el paralelismo psicofísico y, por tanto, adoptaba el punto de vista de que las fuerzas no físicas operaban en la mente, lo que hacía imposible predecir los pensamiento fisiológicamente.)

[31] Véase Spinoza, *Ethics,* primero y último párrafos de la parte III y el estudio de los espíritus animales en el prefacio a la parte V.

[32] Esta «idea 'idea'» plenamente desarrollada (ligeramente presupuesta por Hume en el pasaje citado en la nota 18) es la que constituyó el objeto de la desesperada protesta de Reid; en esta protesta fue precedido por Arnauld y sucedido en los siglos posteriores por T. H. Green y por John Austin. John Yolton me ha señalado un párrafo de Arnauld, en *Des vraies et des fausses idées (Oeuvres,* París y Lausanne, 1780, vol. 39, pág. 190), que utiliza las imágenes del espejo (que se remontan a Platón, *República,* 510a) que yo vengo considerando como el pecado original de la epistemología:

> Como todos los hombres fueron antes niños, y como en esa etapa sólo se ocupaban de sus cuerpos y de las cosas que encontraban sus sentido, han pasado mucho tiempo sin conocer ninguna otra visión *(vue)* que la corpórea, que atribuyen a sus ojos. No podía dejar de fijarse en dos hechos. El primero es que es necesario que, para ver el objeto, éste ha de estar ante sus ojos. Esto es lo que llaman presencia *(presénce)* y hace que consideren que la presencia del objeto es necesaria para la visión. El segundo hecho es que algunas veces vemos cosas visibles en los espejos, o en el agua, o en otras cosas que los representan. Por eso creen, equivo-

Hasta para el mismo Descartes, el tema de la relación entre el cuerpo y el alma no era competencia de la filosofía; ésta, por así decirlo, se había colocado por encima de la sabiduría práctica buscada por los hombres de la antigüedad y había adquirido carácter profesional, casi tan profesional como las matemáticas, cuyo contenido simbolizaba la indubitabilidad característica de la mente. «Sólo en la vida diaria y el discurso corriente, renunciando a meditar y estudiar las materias que excitan la imaginación, se aprende a captar *(concevoir)* la unión del cuerpo y el alma... esa unión que todos experimentan sin filosofar» [33]. El cambio cartesiano de la mente-como-razón a la mente-como-escenario-interno no fue tanto el triunfo del arrogante sujeto individual liberado de las trabas escolásticas cuanto el triunfo de la búsqueda de certeza sobre la búsqueda de sabiduría. De ese momento en adelante, quedaba abierto el camino para que los filósofos consiguieran el rigor del matemático o del físico matemático, o explicaran la apariencia de rigor en estos campos, en vez de tratar de ayudar al hombre a conseguir la paz mental. La filosofía se ocupó de la ciencia, más que de la vida, y su centro fue la epistemología.

6. El dualismo y la «sustancia de la mente»

Podría resumir el resultado de la sección precedente diciendo que la idea de la «separación entre mente y cuerpo» significa cosas diferentes, y se prueba con argumentos filosóficos distintos, antes y después de Descartes. La epistemología hilemórfica que consideraba que la comprensión de los universales consistía en la reproducción en el propio intelecto de lo que la rana reproducía en su carne estaba siendo reemplazada, gracias al ascenso de la física matemática, por un marco de referencia ley-evento que explicaba la raneidad

cadamente, que entonces no ven las cosas mismas, sino únicamente sus imágenes.

Compárese lo que dice Austin sobre «el uso que hacen los filósofos de 'percibir directamente'» *(Sense and Sensibilia* [Oxford, 1962], pág. 19) y sobre las imágenes del espejo *(ibid.,* págs. 31, 50). Hay un estudio interesante de las versiones post-cartesianas de lo que habría querido decir Descartes con «idea» (incluyendo el último esfuerzo de Arnauld por considerar las ideas como actos, a la manera de Brentano, Husserl y G. E. Moore), en Robert McRae, «'Idea' as a Philosophical term in the 17th Century», *Journal of the History of Ideas,* 26 (1965), 175-190.

[33] Carta a la princesa Isabel, 28 de junio de 1643 (edición Alquié, III, 45), citada en van Peursen, *Body, Soul, Spirit,* pág. 25.

como si fuera, quizá, más que una mera esencia «nominal». Por eso, la idea de la razón en cuanto facultad para captar los universales era algo que no se podía utilizar en una premisa que probaba la diferencia entre mente y cuerpo. La idea que podría determinar lo que pudiera «tener una existencia distinta del cuerpo» sería aquella que consiguiera trazar una línea divisoria entre los retortijones experimentados en el estómago y la sensación asociada en la mente.

Ya he señalado que el único criterio capaz de trazar esta línea es la indubitabilidad —esa proximidad al Ojo Interno que permite a Descartes decir (en una frase que habría asombrado a Isabela y a la antigüedad) que «nada conoce la mente con tanta facilidad como a sí misma» [34]. Pero esto puede resultar extraño, pues parece que el candidato más claro a esta distinción es la no-espacialidad. Descartes insiste una y otra vez en que no podemos separar la mente de la «sustancia extensa», considerándola así como sustancia no-extensa. Además, la primera refutación, y la más inspirada en el sentido común, presentada a los filósofos contemporáneos que indican que los dolores podrían identificarse con los procesos cerebrales está tomada directamente de Descartes: a saber, que los dolores que se producen «en» los miembros amputados son no-espaciales —donde la argumentación es que si tuvieran *alguna* localización espacial estarían en un brazo, pero como no hay brazo, tienen que ser de una naturaleza ontológicamente muy diferente. Los filósofos siguen insistiendo en que *«no tiene ningún sentido* localizar la producción de un pensamiento en algún lugar del cuerpo» [35], y suelen atribuir esta idea a Descartes. Pero, como he tratado de demostrar en la sección 2, difícilmente pensaríamos que un pensamiento o un dolor es una *cosa* (un particular distinto de una persona, y no un estado de una persona) que no era localizable a no ser que tuviéramos ya la idea de una sustancia no extensa de la que podría ser una parte. Ninguna intuición de que los dolores y pensamientos sean no-espaciales precede en el tiempo a (ni puede

[34] Este pasaje se sitúa en una espléndida «conclusión errónea» que sigue al ejemplo del trozo de cera en *Meditación II*: como ni siquiera los cuerpos son conocidos «proprie» por el sentido o la imaginación, sino únicamente por el intelecto, es claro que *nihil facilius aut evidentius mea mente posse a me percipi* (Alquié, II, 192; Haldane y Ross, I, 157). El argumento se basa en la confusión entre el *cogito* como prueba de mi existencia y como aislamiento de mi esencia.

[35] Jerome Shaffer, *Philosophy of Mind* (Englewood Cliffs, N. J., 1968), página 48; cfr. Norman Malcolm, «Scientific Materialism and the Identity Theory», *Dialogue 3* (1964), 115-125.

servir de base para un argumento en favor de) la idea cartesiana de la mente en cuanto sustancia distinta (no-espacial). No obstante, queda algo por decir sobre la forma en que entró en la filosofía la idea de «sustancia no-espacial» y, por tanto, de «aquello de que está hecha la mente», y, en consecuencia, sobre las razones por las que la filosofía contemporánea de la mente habla sobre *dolores* y *creencias* y no de *personas que tienen* dolores o creencias. Examinando estos puntos quedará más clara, espero yo, la enorme diferencia entre el dualismo *cartesiano* y el «dualismo» de las obras contemporáneas.

Debemos tener presente que la sustancia no-espacial que Kant y Strawson rechazan por incoherente fue una idea del siglo XVII, y que es un tópico de la historia intelectual que en aquel siglo ocurrieron cosas muy extrañas con la idea de «sustancia». Para Aristóteles, y también para Santo Tomás, el paradigma de una sustancia era un hombre o una rana individual. Las partes separadas de un hombre o de una rana eran, como un manojo de hierba o un cubo de agua, casos dudosos —eran «capaces de existir separadamente» en un sentido (separación espacial), pero no tenían la unidad funcional o «naturaleza» que deben tener las sustancias propiamente dichas. Aristóteles, cuando se ocupaba de estos casos problemáticos, solía descartarlos como «meras potencias» —ni accidentes, como el color de la rana, ni realidades propiamente dichas, como la misma rana que vive y salta [36]. Descartes hace como que utiliza «sustancia distinta» en el sentido clásico de «capaz de existencia separada», pero no se refiere *ni* a la separación *ni* a la unidad funcional [37]. Se refiere a algo parecido a «capaz de hacer que desapa-

[36] Cfr. *Metaphysics,* 1040b, 5-10, sobre «montones». He estudiado la tensión entre los criterios de Aristóteles para ser una sustancia —«existencia separada» y «unidad»— en «Genus as Matter», en *Exegesis and Argument: Studies in Greek Philosophy in Honor of Gregory Vlastos,* ed. Edward N. Lee *et al.* (Dordrecht, 1973).

[37] Descartes piensa que una mano humana, separada o no, es un ejemplo perfecto de sustancia. Cfr. *Fourth Responses* (Alquié, II, 663, Haldane y Ross, II, 99), donde dice que el sentido en que una mano es una «sustancia incompleta» carece de importancia y sólo es cuestión de «no formar un todo distinto de cualquier otra cosa» *(en un autre sens on les peut appeler incomplètes, non qu'elles aient rien d'incomplet en tant qu'elles sont des substances, mais seulement en tant qu'elles se rapportent à quelqu-autre substance avec laquelle elles composent un tout par soi et distinct de tout autre).* Hay, sin embargo, una tensión en Descartes entre la opinión de que cualquier cosa material (una mano, una mota de polvo) es una sustancia y la idea (clara en Spinoza) de que estas cosas son sólo modos de una sustancia más amplia (por ejemplo, la materia-como-un-todo).

rezca todo lo demás [o se excluya mentalmente] y a pesar de ello seguir estando ahí»[38]. Esta definición de «capacidad de existencia separada» se aplica al Uno, a las ideas platónicas y a los motores no movidos de Aristóteles, pero casi a nada más. Dada esta definición, no es de extrañar que resulte que sólo hay, como máximo, tres sustancias: Dios, la mente y la materia. Tampoco debe sorprendernos que Malebranche y Berkeley comiencen a tener dudas sobre el tercer candidato, y Spinoza sobre el segundo y el tercero. A Aristóteles no se le habría ocurrido pensar que las ranas, estrellas y hombres no eran más que otros tantos accidentes de una gran sustancia por la única razón de que si imaginamos que se aniquilan todos los demás cuerpos del mundo (por ejemplo, la tierra y el aire), la rana y el hombre difícilmente podrían sobrevivir. Pero fue precisamente esta noción de una gran sustancia la que resultó necesaria para ofrecer un «fundamento filosófico» a la mecánica de Galileo y desprestigiar las tradicionales explicaciones hilemórficas[39]. Cuando la materia-en-cuanto-unión-de-todos-los-átomos (o vértices) sustituyó a la materia-en-cuanto-potencialidad, fue promovida al rango de sustancia (que absorbía en sí misma todas las antiguas sustancias aristotélicas no humanas) y dejaba únicamente la «pura realidad» de Aristóteles (el νοῦς que es los Motores No Movidos y quizá no se distinguía del νοῦς «separable» de los hombres individuales) como posible rival en ese rango[40].

[38] Esta definición plantea problemas a Descartes, pues sugiere el punto de vista spinozanio de que Dios, de quien hay que pensar que depende todo lo demás, es la única sustancia. El tema lo estudia L. J. Beck en *The Metaphysics of Descartes* (Oxford, 1965). En la página 110 dice: «La aparente incoherencia de usar la palabra *substantia*, o incluso *res*, para connotar al yo del *Cogito* se debe, en bastante medida, al intento de poner vino nuevo en odres viejos, de expresar la doctrina cartesiana en el lenguaje técnico de las escuelas.»

[39] Cfr. E. A. Burtt, *Metaphysical Foundations of Modern Physical Science* (Garden City, N. Y., 1955), capítulo IV. En la página 117 Burtt dice: «El hecho es, y esto tiene una importancia fundamental para todo nuestro estudio, que *el criterio real de Descartes no es la permanencia, sino la posibilidad de manipulación matemática;* en su caso, como en el de Galileo, todo el curso de su pensamiento, desde sus estudios de adolescencia en adelante, le habían llevado a la idea de que sólo conocemos los objetos en términos matemáticos.» La distinción resultante entre cualidades primarias y secundarias proporciona el motivo para considerar las sustancias aristotélicas paradigmáticas como modos de la *res extensa*, nada más.

[40] Cfr. *Fourth Responses* (Alquié, II, 662; Haldane y Ross, II, 98), donde Descartes dice que *«concevoir pleinement»* y *«concevoir que c'est une chose complète»* son sinónimos, lo cual, cree él, sirve para explicar cómo comprendemos que el alma y el cuerpo son dos sustancias.

Los hombres de nuestro tiempo, herederos de la distinción cartesiana entre mente y materia, hemos perdido contacto con la idea de «sustancia» en su definición del siglo XVII. La idea de la existencia *a se* nunca fue inteligible para la gente común, y Kant consiguió hacerla ininteligible hasta para los filósofos profesionales. Por eso, cuando *nosotros* asentimos a la afirmación de que hay una distinción evidente entre una categoría de cosas que pueden existir en el espacio y otra categoría de cosas que no pueden, no estamos asintiendo a la afirmación cartesiana de que la materia y la mente son entidades distintas «que no dependen de ninguna otra cosa para su existencia». Muchos filósofos contemporáneos que están de acuerdo en que es absurdo hablar de la localización de un dolor o un pensamiento insisten, no obstante, y a pesar de Descartes, en que es inimaginable una corriente de la conciencia sin un cuerpo. Estos filósofos se conforman con hablar de las entidades mentales como estados de las personas y no como «trocitos de una sustancia fantasmagórica», y admiten que la no-localizabilidad es un signo de la categoría adjetiva de los estados más que de la estructura peculiar de algunos particulares. Si no es posible localizar con toda precisión en el espacio el tipo, personalidad, peso, hilaridad o encanto de una persona, ¿por qué han de poder serlo sus creencias y deseos? Parece plausible decir que la idea de Descartes fue simplemente un reconocimiento de la diferencia que hay entre las partes de las personas o los estados de esas partes (por ejemplo, retortijones del estómago), por un lado, y ciertos estados de toda la persona, por el otro, formulado engañosamente en un vocabulario escolástico corrompido como una distinción de «sustancia».

Esta explicación de lo que quiere decir que la mente es no espacial constituye una forma adecuada de afirmar y eliminar, simultáneamente, *un* problema mente-cuerpo. Son pocas las personas que se preocupan por la existencia de un abismo ontológico entre lo que significa con los nombres y lo que se significa con los adjetivos. Sin embargo, como la mayoría de las soluciones de tipo conductista al problema mente-cuerpo, ésta tiene dificultades con los pensamientos y sensaciones primarias —hechos en oposición a disposiciones. Es fácil considerar las creencias y deseos y humores como (en expresión de Ryle) «rasgos del entendimiento y el carácter» que no necesitan ningún medio no material en cuanto substrato sino únicamente al hombre mismo. Cuesta más pensar de esta manera en las sensaciones primarias, imágenes mentales y pensamientos [41].

[41] La afirmación aparece de varias maneras en Hampshire, Austin y Ayer, en sus respectivas revisiones de *The Concept of Mind,* reimpresas en *Ryle:*

Nos hacen pensar en un curso inmaterial de la conciencia que corre invisible e intangiblemente por entre los intersticios del cerebro, quizá —pues parece muy natural considerarlos como cosas más que como estados de las cosas. Por eso, los filósofos contemporáneos, volviendo a la noción aristotélica y vulgar de «cosa» en vez de a la rebuscada y extravagante noción cartesiana de «sustancia», se inclinan a dividir la diferencia entre Aristóteles y Descartes. Es decir, piensan que Aristóteles descuidó ciertos particulares —por ejemplo, los dolores y sensaciones primarias— mientras que Descartes consideró inútilmente que eran accidentes de una gran sustancia no extensa, lo mismo que pensaba que las ranas y los átomos eran configuraciones accidentales de una gran sustancia extensa llamada Materia. Esto hace posible que los filósofos actuales hablen de entidades mentales sin alma, y, por tanto, sin que parezca que están perseguidos por el Hombre Invisible e Intangible de la creencia religiosa (idea que creyeron ver en Descartes —no sin cierta ayuda por parte del mismo).

Este dualismo basado en la «existencia separada del cuerpo» —una cuarta clase— es muy diferente del dualismo entre una persona y su fantasma, o entre una persona y su entendimiento pasivo aristotélico, o entre *res cogitans* y *res extensa*. Pero es también un dualismo parcial —tan parcial, en formas iguales y opuestas, como el de los antiguos. Mientras que los antiguos consideraron que sólo «existían separadamente» la parte de la sustancia no extensa de Descartes que comprendía los universales, los dualistas contemporáneos (concediendo a Ryle las creencias, deseos y cosas semejantes como formas de hablar de disposiciones) sólo admiten la «existencia separada» de los candidatos a la categoría mental que tienen carácter de hecho. Mientras los tomistas, por ejemplo, acusan a Descartes de haber realizado la tarea absurda de dotar al sentido de la inmaterialidad que es rasgo prerrogativo de la razón, los dualistas contemporáneos le acusan de haber realizado la tarea absurda de haber dotado al conocimiento matemático y a las decisiones sobre la conducta del carácter de cosa inmaterial que pertenece a los dolores, post-imágenes y pensamientos que ocurren. Para los

---

A Collection of Critical Essays, ed. O. P. Wood y George Pitcher (Garden City, N. Y., 1970). Sobre las formas de ampliar los métodos de Ryle a las percepciones y a las sensaciones primarias asociadas, véase Pitcher, *A Theory of Perception*. Sobre las percepciones en cuanto disposiciones a creer, y la discusión del «materialismo adverbial», véase James Cornman, *Materialism and Sensations* (New Haven, 1971). Véase también Richard Rorty, «Incorrigibility as the Mark of the Mental», *Journal of Philosophy*, 67 (1970), 406-409.

antiguos, la mente era, con toda evidencia, capaz de existencia separada cuando contemplaba lo incambiable y era ella misma incambiable. Para los modernos, cuando más evidente es que tiene esa capacidad es cuando es una colección floreciente y ruidosa de sensaciones primarias[42]. Sea quién sea quien está en lo cierto, es claro que ni los antiguos ni los modernos comparten la «percepción clara y distinta» cartesiana de la separabilidad de *todos* los elementos que amontonó bajo el epígrafe «pensar».

El único perfeccionamiento de Descartes en relación con la idea homérica del Hombre Invisible e Intangible fue privar al intruso de forma humanoide. Al conseguir con ello que los posibles intrusos entre los cuerpos fueran menos fácilmente identificables, los hizo más filosóficos. Eran más filosóficos en cuanto que, como el νοῦς de Aristóteles y la Esencia de Vidrio de Isabela, no eran oscuros homúnculos, sino más bien realidades esencialmente imposibles de reproducir gráficamente. Como ocuparse de asuntos filosóficos era ocuparse de lo que el ojo no puede ver ni el oído oír, se pensó que tanto la sustancia no extensa del siglo XVII como las sensaciones y pensamientos no localizables contemporáneos eran más filosóficamente respetables que los fantasmas por cuya paz rezan los creyentes. Pero los filósofos contemporáneos, tras poner al día a Descartes, pueden ser dualistas sin que su dualismo implique la más mínima diferencia para ningún interés o preocupación del hombre, sin interferir con la ciencia o servir de apoyo a la religión. En la medida en que el dualismo se reduce a la mera insistencia en que los dolores y pensamientos no tienen lugar, no hay nada en absoluto que dependa de la distinción entre mente y cuerpo.

Ruego al lector me permita recordar el curso seguido en este capítulo. En las secciones 1-2 he señalado que no podríamos ver ningún sentido en la noción de «entidades mentales» en cuanto género ontológico distinto sin invocar la idea de «entidades fenoménicas» como los dolores, entidades cuyo ser se agotaba en la única propiedad, por ejemplo, de ser doloroso. He explicado cómo el verdadero problema no era abjurar de estos universales hipostatizados sino explicar por qué se habían tomado en serio y cómo llegaron a parecer relevantes en las discusiones sobre la naturaleza de la personalidad y la razón. Espero que las secciones 3-6 hayan

[42] «Mientras que, para Platón, los pensamientos racionales eran las actividades paradigmáticas del alma, los hechos tan bajos como picores, dolores de muelas y punzadas están ahora entre los típicamente mencionados en la discusión filosófica como hechos mentales» (Jaegwon Kim, «Materialism and the Criteria of the Mental», *Synthese,* 22 [1971], 336).

ofrecido una idea de cómo pienso que se puede dar respuesta a estas cuestiones históricas (aunque soy dolorosamente consciente de las lagunas que aparecen en la historia tal como la he contado). Mi respuesta a la pregunta «¿Por qué tendemos a agrupar lo intencional y lo fenoménico con el término «lo mental»? es que Descartes utilizó la idea de lo «conocido incorregiblemente» para salvar el abismo existente entre ambos. Por eso, ahora debo explicar con mayor profundidad mi propia concepción anti-cartesiana, wittgensteiniana, de la naturaleza de «nuestro acceso privilegiado a lo mental». Así pues, en el siguiente capítulo, dejo de lado la personalidad y la razón y examino casi exclusivamente la conciencia. Trataré de demostrar que el «problema de la conciencia», supuestamente metafísico, no es ni más ni menos que el «problema del acceso privilegiado» —problema epistemológico— y que una vez que se consiga ver esto pierden interés las cuestiones sobre dualismo versus materialismo.

CAPÍTULO II

Personas sin mente

1. LOS ANTÍPODAS

Muy lejos, en el otro extremo de nuestra galaxia, había un planeta en que vivían seres como nosotros —bípedos implumes que construían casas y bombas, y escribían poemas y programas de computador. Estos seres no sabían que tenían mente. Tenían ideas como «querer» o «intentar» y «creer que» y «encontrarse fatal» y «encontrarse muy bien». Pero no tenían la menor idea de que éstas significaran estados *mentales* —estados de un tipo peculiar y distinto— totalmente diferentes de «sentarse», «tener un catarro» y «estar excitado sexualmente». Aunque utilizaban las ideas de creer y saber y querer y estar triste, tanto para hablar de sus animales y robots como de sí mismos, consideraban que los animales o robots no estaban incluidos en lo que se quería decir cuando decían: «Todos nosotros creemos...» o «Nunca hacemos cosas así...» Es decir, sólo trataban como *personas* a los miembros de su especie. Pero no explicaban la diferencia entre personas y no-personas recurriendo a nociones como «mente», «conciencia», «espíritu» ni nada semejante. No la *explicaban* de ninguna manera; simplemente la trataban como la diferencia entre «nosotros» y todo lo demás. Creían en la inmortalidad para ellos mismos, y algunos creían que ésta sería compartida por los animales domésticos o los robots, o ambos. Pero esta inmortalidad no implicaba la idea de un «alma» que se separara del cuerpo. Era sencillamente una resurrección corporal seguida por un movimiento misterioso e instantáneo a lo que ellos llamaban «un lugar en los cielos», en el caso de las buenas personas, y a una especie de caverna, bajo la superficie del planeta, para los malos. Sus filósofos se ocupaban principalmente de cuatro temas: la na-

turaleza del Ser, las pruebas de la existencia de un Ser Benévolo y Omnipotente que podría tomar las medidas necesarias para la resurrección, los problemas que surgían en el discurso sobre objetos inexistentes y la reconciliación de las intuiciones morales contradictorias. Pero estos filósofos no habían formulado el problema del sujeto y el objeto, ni el de la mente y la materia. Había una tradición de escepticismo pirroniano, pero no se conocía el «velo de ideas» de Locke, pues tampoco se conocía la noción de «idea» o «percepción» o «representación mental». Algunos de sus filósofos pronosticaban que las creencias sobre la inmortalidad que habían ocupado un puesto central en periodos anteriores de la historia, y que todavía eran mantenidos por todos los que no formaban parte de la «intelligentsia», se verían reemplazadas algún día por una cultura «positivista» purgada de todas las supersticiones (pero estos filósofos no mencionaban para nada una etapa «metafísica» intermedia).

En la mayoría de los aspectos, el lenguaje, la vida, la tecnología y la filosofía de esta raza eran muy semejantes a los nuestros. Pero había una diferencia importante. La neurología y la bioquímica habían sido las primeras disciplinas en que se habían conseguido avances tecnológicos, y una gran parte de la conversación de estos hombres hacía referencia al estado de sus nervios. Cuando sus niños se acercaban hacia la estufa, las madres gritaban: «Van a estimular sus fibras-C.» Cuando se les presentaban ilusiones visuales ingeniosas, decían: « ¡Qué extraño! Hace parpadear el haz neurónico G-14, pero cuando lo miro desde un lado puedo ver que no es un rectángulo rojo, ni mucho menos.» Su conocimiento de la fisiología era tal que cada oración bien construida que se utilizaba podría relacionarse fácilmente con un estado neural fácilmente identificable. Este estado ocurría siempre que alguien pronunciaba, o tenía intención de pronunciar, u oía, la oración. Este estado ocurría también algunas veces en la soledad y hablaban de estas ocasiones con comentarios como «Me encontré de repente en un estado S-296, y saqué las botellas de leche». Algunas veces dirían cosas como ésta: «Parecía un elefante, pero entonces se me ocurrió que los elefantes no viven en este continente, por lo que me di cuenta de que tenía que ser un mastodonte.» Pero también dirían algunas veces, en circunstancias idénticas, algo parecido a: «Tuve G-412 junto con F-11, pero entonces tuve S-147, por lo que me di cuenta de que tenía que ser un mastodonte.» Pensaban en los mastodontes y botellas de leche como objetos de creencias y deseos, y como causantes de ciertos procesos neurales. Consideraban estos procesos

neurales en interacción causal con las creencias y deseos —exactamente de la misma manera que los mastodontes y las botellas de leche. Ciertos procesos neurales podían ser auto-inducidos deliberadamente, y algunas personas eran más hábiles que otras para inducir en sí mismos ciertos estados neurales. Otras tenían habilidad para detectar ciertos estados especiales que la mayoría de las personas no podían reconocer en sí mismas.

A mediados del siglo XXI llegó hasta este planeta una expedición procedente de la Tierra. En ella iban filósofos, así como representantes de todas las demás disciplinas cultas. Los filósofos pensaban que lo más interesante de los nativos era su carencia del concepto de mente. Bromeaban entre sí diciendo que habían venido a parar a un lugar poblado de materialistas, y proponían para aquel planeta el nombre de Antípodas —haciendo referencia a una escuela casi olvidada de filósofos, con centro en Australia y Nueva Zelanda, que en el siglo anterior habían intentado realizar una de las muchas rebeliones inútiles contra el dualismo cartesiano en la historia de la filosofía terrestre. El nuevo nombre cuajó, y con el tiempo la nueva raza de seres inteligentes fue conocida con el nombre de Antípodas. Los neurólogos y bioquímicos terrestres quedaron fascinados ante la abundancia de conocimientos que los antípodas tenían en su campo. Como la conversación técnica sobre estos temas se realizaba casi enteramente haciendo referencias casuales a estados neurales, los expertos terrestres llegaron a adquirir la capacidad de informar de sus propios estados neurales (sin inferencia consciente) en vez de hablar de sus pensamientos, percepciones y sensaciones primarias. (Las fisiologías de las dos especies eran, por fortuna, casi idénticas.) Todo iba a las mil maravillas, si exceptuamos las dificultades con que tropezaron los filósofos.

Los filósofos que habían llegado con la expedición estaban divididos, como de costumbre en dos campos enfrentados: los más moderados pensaban que la filosofía debía estar orientada hacia la Significación, y los más duros pensaban que debía estar orientada hacia la Verdad. Los filósofos del primer grupo pensaban que no era realmente problema saber si los Antípodas tenían mente. Defendían que lo que importaba para comprender a los demás seres era comprender su modo de estar-en-el-mundo. Resultó evidente que, cualquiera que fuera el *Existentiale* que estaban utilizando los Antípodas, no incluían ninguno de los que, un siglo antes, Heidegger había criticado como «subjetivistas». Toda la idea de «sujeto epistemológico» o de la persona en cuanto espíritu quedaba totalmente excluida de sus auto-descripciones, y de sus filosofías. Algunos de

los filósofos del grupo más moderado consideraban que esto demostraba que los Antípodas no habían conseguido liberarse de la Naturaleza para entrar en el Espíritu, o, dicho con más suavidad, no habían avanzado de la Conciencia a la Auto-Conciencia. Estos filósofos se convirtieron en pregoneros de la interioridad, empeñados en hacer que los Antípodas atravesaran una línea invisible y penetraran en el Reino del Espíritu. Otros, sin embargo, pensaban que los Antípodas estaban en posesión de la valiosa comprensión de la unión de πόλεμος y λόγος que entre los Terrestres Occidentales se había perdido como consecuencia de la asimilación platónica de ουσία y ιδέα. El que los Antípodas no llegaron a captar la noción de mente, en opinión de este grupo de filósofos, demostraba su proximidad con el Ser y su libertad de las tentaciones a las que el pensamiento terrestre había sucumbido hacía mucho tiempo. En el enfrentamiento entre los dos puntos de vista, ambos dentro de la corriente más moderada, las discusiones no llegaban a resultados concluyentes. Los Antípodas no podían prestar demasiada ayuda, pues les costaba mucho traducir las lecturas básicas necesarias para apreciar el problema —*Teeteto* de Platón, las *Meditaciones* de Descartes, el *Tratado* de Hume, la *Crítica de la razón pura* de Kant, la *Fenomenología* de Hegel, *Individuos* de Strawson, etc.

Los filósofos de la línea dura, como de costumbre, encontraron un tema de discusión que resultaba mucho más claro y preciso. No les interesaba lo que pudieran pensar los Antípodas de sí mismos, sino que se centraban en la pregunta: ¿Tienen mente, en realidad? En su afición a la precisión, reducían esta pregunta a: ¿Tienen realmente sensaciones? Se pensaba que si se aclaraba si tenían, por ejemplo, sensaciones de dolor, además de estimulación de las fibras-C, cuando tocaban una estufa encendida, todo lo demás sería coser y cantar. Era claro que los Antípodas tenían las mismas disposiciones conductuales hacia las estufas encendidas, los calambres musculares, la tortura y cosas semejantes, que los seres humanos. Aborrecían la estimulación de sus fibras-C. Pero los filósofos de la línea dura se preguntaban: ¿Contiene su experiencia las mismas propiedades fenoménicas que la nuestra? ¿Resulta dolorosa la estimulación de las fibras-C? ¿O produce una sensación diferente, igualmente terrible? ¿O no se produce ninguna sensación en absoluto? A estos filósofos no les extrañaba que los Antípodas pudieran presentar informes no-inferenciales de sus propios estados neutrales, pues se sabía desde hacía tiempo que los psicofisiólogos podían enseñar a los sujetos humanos a informar de los ritmos alfa, así como de otros distintos estados corticales fisiológicamente descriptibles.

Pero se sentían desconcertados ante la pregunta: ¿Detecta algunas propiedades fenoménicas el Antípoda que dice: «Ya están otra vez mis fibras-C, ya sabes, las que se disparan cuando te quemas o te das un golpe o te sacan una muela. Es terrible.»?

Se insinuó que la pregunta sólo podía recibir respuesta de forma experimental, y de esta forma llegaron a un acuerdo con los neurólogos para que uno de ellos quedara conectado con un Antípoda voluntario de forma que se estableciera un intercambio de corriente entre las distintas regiones de los dos cerebros. Se pensaba que esto permitiría a los filósofos asegurarse de que los Antípodas no tenían un espectro invertido ni ninguna otra cosa que pudiera confundir al tejido. Sin embargo, la realidad fue que el experimento no produjo ningún resultado interesante. La dificultad estaba en que cuando el centro del habla del Antípoda recibía una entrada de las fibras-C del cerebro terrestre siempre se limitaba a hablar de sus fibras-C, mientras que cuando el centro de habla del Terrestre controlaba la situación siempre se limitaba a hablar de dolor. Cuando se preguntaba al centro de habla del Antípoda qué sensación tenían las fibras-C, decía que no entendía la idea de «sensación», pero que resultaba terrible que se produjera la estimulación de las fibras-C. Lo mismo ocurría con las preguntas sobre los espectros invertidos y otras cualidades perceptivas. Cuando se les indicaba que dijeran en voz alta los colores de un gráfico, ambos centros de habla decían los nombres normales de los colores en el mismo orden. Pero el centro de habla del Antípoda podía enumerar también los distintos paquetes neurónicos activados por cada parte del gráfico (independiente de la corteza visual a que estuviera acoplado). Cuando se preguntaba al centro de habla del Terrestre cómo eran los colores cuando se transmitían a la corteza visual del Antípoda, decía que igual que siempre.

Parecía que el experimento no había servido para nada. No se había aclarado si los Antípodas tenían dolores. Seguía siendo igual de oscuro que antes si tenían una o dos sensaciones primarias cuando se proyectaba sobre sus retinas luz de color añil (una de añil y otra de estado neural C-692) —o si tenían ninguna sensación primaria en absoluto. Se preguntaba repetidamente a los Antípodas cómo sabían que era añil. Respondían que veían que lo era. Cuando se les preguntaba cómo sabían que estaban en C-692, decían que lo «sabían», sin más. Cuando se les indicaba que podrían haber inferido inconscientemente que era añil basándose en la sensación C-692, parecían incapaces de entender qué era una inferencia inconsciente o qué era una «sensación». Cuando se les señalaba que

podrían haber hecho la misma inferencia del hecho de que estaban en el estado C-692 basándose en la sensación primaria de añil, se quedaban igualmente desconcertados. Cuando se les preguntaba si el estado neural parecía añil, respondían que no —era añil la *luz*— y que el que formulaba la pregunta debía estar cometiendo algún error de categoría. Cuando se les preguntaba si podían imaginarse que tenían C-692 sin ver añil, decían que no. Cuando se les preguntaba si era una verdad conceptual o una generalización empírica que estas dos experiencias iban juntas, respondían que no estaban seguros de como se podía saber la diferencia. Cuando se les preguntaba si podían equivocarse cuando decían que veían añil, respondían que sí, pero que no podían equivocarse cuando decían que les parecía que veían añil. Cuando se les preguntaba si podían equivocarse al decir que estaban en el estado C-692, respondían exactamente igual. Finalmente, una hábil dialéctica filosófica les llevó a darse cuenta de que lo que no podían imaginar era que pareciera que veían añil y que no pareciera que estaban en un estado C-692. Pero con esto parecía que se conseguía avanzar en relación con las preguntas «¿Sensaciones primarias»? «¿Dos sensaciones primarias o una»? «¿Dos referentes o un referente bajo dos descripciones?» Tampoco se progresaba en la pregunta sobre la forma en que se les presentaban las fibras-C estimuladas. Cuando se les preguntaba si podían equivocarse al pensar que se habían estimulado sus fibras-C, respondían que sí —pero que no podían imaginar que estuvieran equivocados al hablar de si parecía que sus fibras-C habían sido estimuladas.

En este contexto, a alguien se le ocurrió preguntar si podían detectar el estado neural que era concomitante de «parecer que se habían estimulado las fibras-C». Los Antípodas respondieron que estaba, por supuesto, el estado T-435, que era el concomitante neural constante de la alocución de la oración «Parece que se han estimulado mis fibras-C», el estado T-497 que iba con «Tengo la impresión de que se han estimulado mis fibras-C», el estado T-293 que acompañaba a «¡Estimulación de làs fibras-C!» y varios otros estados neurales que eran concomitantes de varias otras oraciones aproximadamente sinónimas —pero que no había más estados neurales de que tuvieran conocimiento, admás de éstos. Los casos en que los Antípodas tenían T-435 y no tenían estimulación de las fibras-C incluían aquellos en que, por ejemplo, se les sujetaba a algo que les habían dicho, engañosamente, que era un instrumento de tortura, se conectaba llamativamente un mando, y no se hacía nada más.

La discusión entre los filósofos se centró entonces en el tema: ¿Podrían equivocarse los Antípodas en relación con la serie-T de estados neurales (los que eran concomitantes de entender o pronunciar oraciones)? ¿Podría parecerles que tenían T-435 sin tenerlo realmente? Sí, decían los Antípodas, los cerebroscopios indicaban que eso era algo que ocurría de vez en cuando. ¿Había alguna explicación de los casos en que ocurría —algún patrón de los mismos? No, parecía que no. Lo único que pasaba era que de vez en cuando se producía una de esas cosas raras. La neurofisiología no había sido todavía capaz de encontrar otro tipo de estado neural, fuera de la serie-T, que fuera concomitante de estas ilusiones misteriosas, como tampoco lo había conseguido en el caso de ciertas ilusiones perceptivas, pero quizá lo consiguiera algún día.

Esta respuesta dejaba a los filósofos con sus problemas ante la pregunta de si los Antípodas tenían sensaciones de dolor, o cualquier otra cosa. Parecía que los antípodas no eran incorregibles en nada, excepto en cómo les parecían las cosas. Pero no era claro que «cómo les parecían las cosas» fuera cuestión de las sensaciones primarias que tenían, en oposición a lo que se inclinaban a decir. Si tenían la sensación primaria del dolor, entonces tenían mente. Pero una sensación primaria es (o tiene) una propiedad fenoménica —una propiedad que no se puede tener la ilusión de tener (porque, por así decir, tener la ilusión de tenerla es tener*la*). La diferencia entre dolores y estimulación de las fibras-C estaba en que se podía tener la ilusión de que se habían estimulado las fibras-C (por ejemplo, se podría tener T-435) sin que éstas hubieran sido estimuladas, pero no se podía tener la ilusión de dolor sin tener dolor. En lo único en que no se podían equivocar los Antípodas era en cómo les parecían las cosas. Pero el hecho de que no era posible que «solamente pareciera que hacían que pareciera que...» no tenía ningún interés para determinar si tenían mente. El hecho de que «parece que parece...» sea una expresión que no se utiliza es un dato sobre la idea de «apariencia», no una información de la presencia de «propiedades fenoménicas». Efectivamente, la distinción apariencia-realidad no se basa en la distinción entre representaciones subjetivas y estados de cosas objetivos; es simplemente cuestión de cometer un error, de tener una falsa creencia. Por eso, el dominio que tenían los Antípodas de la primera distinción no servía a los filósofos para saber si debían atribuirles la segunda.

2. PROPIEDADES FENOMÉNICAS

Volviendo al presente, ¿qué *debemos* decir de los Antípodas? Lo primero que hay que hacer, probablemente, es considerar más atentamente la idea de «propiedad fenoménica» y en especial la diferencia entre aprehender un fenómeno físico en forma engañosa y aprehender un fenómeno mental en forma engañosa. La explicación de esta diferencia presentada por Kripke resume la intuición en que se han basado generalmente los defensores del dualismo, por lo que podemos comenzar nuestro examen tratando de aplicar su terminología:

> Alguien puede estar en la misma situación epistémica en que estaría si hiciera calor, incluso en ausencia de calor, sencillamente teniendo la sensación de calor; e incluso en presencia de calor, puede tener la misma evidencia que tendría en ausencia de calor, sencillamente careciendo de la sensación S. Esta posibilidad no se da en el caso del dolor o de otros fenómenos mentales. Estar en la misma situación epistémica que se obtendría si se tuviera dolor *es* tener dolor; estar en la misma situación epistémica que se daría en ausencia de dolor *es* no tener dolor... El problema está en que la idea de una situación epistémica cualitativamente idéntica a otra en que el observador tuviera una sensación S *es* simplemente aquella en que el observador tenía dicha sensación. Lo mismo se puede decir de la idea de qué es lo que descubre la referencia de un designador rígido [expresión que designa al mismo objeto en todos los mundos posibles en que los designe de alguna manera]. En el caso de la identidad del calor con el movimiento molecular, lo importante a tener en cuenta era que, aunque «calor» sea un designador rígido, la referencia de ese designador estaba determinada por una propiedad accidental del referente, a saber la propiedad de producir en nosotros la sensación S. ... Por otra parte, el dolor no se descubre por una de sus propiedades accidentales, sino más bien por la propiedad de ser el dolor mismo, por su cualidad fenomenológica inmediata. Por eso, el dolor, a diferencia del calor, no sólo se designa rígidamente con «dolor» sino que la referencia del designador está determinada por una propiedad esencial del referente. Así, no es posible decir que aunque el dolor se identifique necesariamente con determinado estado físico, pueda descubrirse un determinado fenómeno de la misma manera que descubrimos el dolor sin estar en correlación con ese estado físico. Si algún

fenómeno se descubre de la misma forma en que se descubre el dolor, entonces ese fenómeno *es* dolor [1].

Estas consideraciones nos indican que la verdadera pregunta es ésta: ¿Descubren los Antípodas los fenómenos mentales por las propiedades accidentales? Si suponemos, de momento, que *tienen* dolores, ¿podrían no dar con la «cualidad fenomenológica inmediata» y fijarse únicamente en el rasgo accidental de la compañía constante de la estimulación de las fibras-C? O, si no pueden, exactamente, *dejar de apreciar* una cualidad fenomenológica inmediata, ¿podría ocurrir que no tuvieran un nombre que atribuirle, y por lo tanto no lograran descubrir la realidad que tiene la cualidad por una propiedad esencial? Dicho de otra manera, como los Antípodas *no* descubren el dolor «exactamente de la misma manera en que nosotros descubrimos el dolor», ¿podemos concluir que lo que tienen, sea lo que sea, *no* es dolor? La relación epistémica de cada uno con sus sensaciones primarias, ¿es necesaria y suficiente para determinar la existencia de la sensación primaria en cuestión? ¿O deberíamos decir que, en realidad, descubren el dolor exactamente igual que nosotros —pues cuando dicen: «¡Ooh! ¡Estimulación de las fibras-C!» sienten exactamente lo mismo que sentimos nosotros cuando decimos. «¡Dolor!» En realidad, quizá, estaban sintiendo dolor y llamaban a esa sensación «estado en que parece que se han estimulado las fibras-C», y están en la misma situación epistémica, por lo que se refiere a que parece que han sido estimuladas sus fibras-C, en que estamos nosotros cuando parece que vemos algo rojo, y en todos los otros estados incorregibles.

Da la impresión de que lo que necesitamos es un criterio muy general para determinar cuándo dos cosas son «realmente» la misma cosa descrita de dos formas diferentes. Parece que no hay en este enigma nada distintivo que lo haga depender de las peculiaridades de lo mental. Si estamos de acuerdo en que lo que cuenta para determinar si los Antípodas tienen sensaciones primarias es la incorregibilidad —la imposibilidad de tener una ilusión de...— el problema general de las descripciones alternativas nos impediría aplicar este criterio y resolver así la cuestión. Se trata de un problema que no va a recibir una solución clara, precisa y de fácil

[1] Saul Kripke, «Naming and Necessity», en *Semantics of Natural Language*, edición Donald Davidson y Gilbert Harman (Dordrecht, 1972), págs. 339-340. Véase una crítica de la exposición del dualismo y materismo hecha por Kripke en Fred Feldman, «Kripke on the Identity Theory», y William Lycan, «Kripke and the Materialists», ambos en *Journal of Philosophy*, 71, (1974), 665-689.

aplicación. No hay nada general que pueda resolver todas las tensiones existentes entre decir:

> Estás hablando de Xs, sin lugar a dudas, pero prácticamente todo lo que dices de ellas es falso

y decir:

> Como prácticamente nada de lo que dices es verdad de Xs, no puedes estar hablando de Xs.

Pero dejemos aparte, por el momento, esta dificultad (volveremos sobre ella en el capítulo seis) y consideremos el aspecto todavía más deprimente de que todo aquel que tratara de formular criterios generales para asimilar o distinguir los referentes de las expresiones necesitaría algunas categorías ontológicas generales —algún procedimiento fuerte, aunque tosco, para esbozar las cosas— simplemente para ponerse en marcha. Sería inútil, en concreto, disponer de una distinción entre entidades mentales y entidades físicas. Pero el problema de los Antípodas cuestiona toda esta distinción. Para ver por qué ocurre esto, supongamos que no hay ningún criterio para los «fenómenos mentales», salvo el epistémico de Kripke[2]. Esta suposición identifica «lo mental» con las sensaciones primarias, pensamientos transitorios e imágenes mentales. Excluye las creencias, humores y cosas semejantes (que, aunque indudablemente «superiores», no forman, sin embargo, parte de nuestra vida interior de la que podemos informar incorregiblemente, y, por lo tanto, no pueden promover la distinción de tipo cartesiano entre dos reinos ontológicos). La suposición equivale, en otras palabras, a la afirmación de que (1) es suficiente para ser un estado mental que la cosa en cuestión sea cognoscible incorregiblemente por su posesor, y (2) no atribuimos literalmente ningún estado no físico (por ejemplo, creencias) a seres que no tengan dichos estados cognoscibles incorregiblemente. (Esto encaja con la práctica de los Antípodas, y con nuestra intuición de que los perros tienen estados no físicos simplemente en virtud de que tienen dolores, mientras que los computadores no los tendrían, aun a pesar de que nos ofrecen verdades nuevas e interesantes). Si se parte de esta suposición, no habrá *nada*

[2] He propuesto una forma matizada de esta suposición en «Incorrigibility as the Mark of the Mental», *Journal of Philosophy,* 67 (1970), 399-424. Véase también Jaegwon Kim, «Materialism and the Criteria of the Mental», *Synthese,* 22 (1972), 323-345, esp. 336-341.

que responder a la pregunta «Cuando informan de que parece que se están activando las fibras-C, ¿están informando de una sensación (quizá la misma sensación de que nosotros informamos diciendo '¡dolor!') o simplemente están produciendo los sonidos que provocan sus neuronas al estar en ciertos estados?» Y si esto es así, dado que el papel desempeñado en nuestras vidas por los informes de sensaciones es el mismo papel desempeñado en las vidas de los Antípodas por los informes de las neuronas, nos encontramos con otra pregunta: Cuando *nosotros* utilizamos el término «dolor», ¿estamos informando de las sensaciones o de las neuronas?

El problema es real, como se comprueba observando las implicaciones de la identidad del papel funcional. Si es cierto que los Antípodas tienen una cultura tan amplia como la nuestra, si son tan intencionales en su discurso y tan auto-conscientemente estéticos en su elección de objetos y personas como nosotros, si su deseo de elevación moral e inmortalidad es igualmente grande, es probable que piensen que el interés de nuestros filósofos en saber si tenemos mente es un poco insignificante. ¿Por qué, se preguntan, es tan grande la diferencia? ¿Por qué, pueden preguntarnos, pensamos *nosotros* que tenemos esas cosas extrañas llamadas «sensación» y «mente»? Ahora que nos han enseñado la micro-neurología, ¿no podemos ver que el hablar de estados mentales no era más que una forma de hablar de neuronas? O, si tenemos realmente algunos estados curiosos además de los neurológicos, ¿son de verdad tan importantes? ¿Constituye realmente la posesión de dichos estados la base para distinguir entre categorías ontológicas?

Estas últimas preguntas ilustran con qué ligereza se toman los Antípodas la controversia que, entre los filósofos terrestres, constituye el reñido debate entre materialistas y epifenomenalistas. Además, el éxito de la neurología Antípoda, no sólo en la explicación y control de la conducta sino también en la creación del vocabulario necesario para la auto-imagen de los Antípodas, demuestra que ninguna de las otras teorías terrestres sobre «la relación entre la mente y el cuerpo» puede tener ni siquiera la posibilidad de triunfar. El paralelismo y el epifenomenalismo sólo pueden diferenciarse con un punto de vista de la causalidad diferente del de Hume —con un punto de vista según el cual haya que descubrir un mecanismo causal que indique qué dirección siguen las líneas causales. Pero nadie, ni el cartesiano más intransigente, se imagina que, cuando tengamos a la vista (como tienen, *ex hypothesi,* los Antípodas) una explicación, molécula-por-molécula, de las neuronas, quedará todavía un lugar donde mirar en busca de nuevos mecanismos causales. (¿A

qué equivaldría «mirar»?) Por eso, aun cuando abandonemos a Hume, no estamos en absoluto en condiciones de ser paralelistas, como no sea por alguna razón a priori según la cual «sabemos sin más» que lo mental es una esfera causal autónoma. En cuanto al interaccionismo, a los Antípodas nunca se les ocurriría negar que las creencias y deseos, por ejemplo, se interaccionen causalmente con las irradiaciones de la retina, los movimientos del brazo, etc. Pero opinan que el hablar de esta interacción no significa unir esferas ontológicas diferentes sino una referencia cómoda (por breve) a la función, más que a la estructura. (Hay tan poco problema, filosóficamente hablando, como en una transacción entre un gobierno y un individuo. No es posible, en el caso de un comentario sobre tal transacción, formular un conjunto de condiciones necesarias y suficientes en términos de quién hizo algo, qué es lo que hizo y a quién lo hizo, en mayor medida que en el caso de los comentarios sobre las creencias causadas por las radiaciones y movimientos causados por las creencias —pero, ¿a quién se le ocurriría pensar qu. sería posible?) La interacción sólo tendría interés si una descarga neural se desviara de su curso por la acción de una sensación primaria, o perdiera parte de su potencia por la misma razón, o algo parecido. Pero los neurólogos antípodas no necesitan tales hipótesis.

Si no hay forma de explicar a los Antípodas nuestros problemas y teorías sobre la mente y el cuerpo —de hacerles ver que es un caso paradigmático de una división ontológica— deberíamos estar dispuestos a enfrentarnos con la posibilidad de que los Antípodas «materialistas» (en oposición a los «epifenomenalistas» más tolerantes) estén en lo cierto: hemos estado hablando de neuronas cuando pensábamos que hablábamos de sensaciones primarias. Sólo ha sido una circunstancia fortuita de nuestro desarrollo cultural que hayamos estado tanto tiempo dando vueltas a lo que sólo estaba haciendo las veces de otra cosa. Es como si, mientras se habían perfeccionado muchas disciplinas sublunares, no hubiéramos desarrollado para nada la astronomía y hubiéramos seguido siendo pre-tolemaicos en nuestras ideas de lo que había por encima de la Luna. Tendríamos, indudablemente, muchas cosas complicadas que decir sobre los agujeros de la bóveda negra, los movimientos de la bóveda en su conjunto, y cosas semejantes —pero una vez que hubiéramos dado con la pista, podríamos redescribir lo que habíamos estado describiendo con relativa facilidad.

Pero, en este punto, hay que hacer frente a una objeción bien conocida. Se expresa en comentarios como éste:

> ... en el caso de los dolores punzantes, no es posible mantener que la micro-imagen sea la imagen real, que las apariencias perceptivas sean sólo una reproducción tosca, pues en este caso estamos tratando de las mismas apariencias perceptivas, que difícilmente pueden ser una tosca reproducción de sí mismas [3].
>
> Está muy bien afirmar que la actividad de las fibras-C de la corteza aparece en forma de nocividad, que el olor a cebolla es la manera en que aparece a un ser humano con un sistema nasal normal la forma de las moléculas de cebolla... De esta manera se considera el dolor, olor o color aprehendidos y, relegándolos a la categoría de apariencia, se los convierte en ontológicamente neutros. Pero nos deja con un conjunto de *pareceres,* actos de aprehensión imperfecta, en los que se captan las propiedades fenoménicas. Por tanto, debemos plantearnos la pregunta: ¿Es posible que las cosas *parezcan que son* de una determinada manera a un sistema meramente material? ¿Hay alguna forma en que los actos de aprehensión imperfecta puedan considerarse como ontológicamente neutrales?
>
> ... La explicación materialista de los hombres reales no puede encontrar un lugar para el hecho de que nuestra aprehensión imperfecta se hace mediante una propiedad fenoménica y no, por ejemplo, mediante creencias que se presentan espontáneamente [4].

Esta objeción, común a Brandt y Campbell, a primera vista parece consistir en que sólo se pueden describir erróneamente las cosas si no se es un «sistema meramente material» —pues dichos sistemas no pueden hacer que las cosas les parezcan distintas de lo que son. Pero esto no puede mantenerse tal cual, pues, como he señalado anteriormente, la distinción entre realidad y apariencia parece ser meramente la distinción entre captar bien las cosas y captarlas mal —distinción que no tenemos ningún problema en hacer en el caso de los robots, servo-mecanismos, etc. Para que la objeción sea plausible debemos decir que, en el presente contexto, «apariencia» es una noción más rica —que ha de explicarse por la noción de la «propiedad fenoménica». Hemos de adoptar un principio parecido al siguiente:

> (P) Siempre que hagamos un informe incorregible sobre un estado de nosotros mismos, tiene que haber una propie-

---

[3] Richard Brandt, «Doubts about the Identity Theory», en *Dimensions of Mind,* ed. Sidney Hook (Nueva York, 1961), pág. 70.

[4] Keith Campbell, *Body and Mind* (Nueva York, 1960), págs. 106-107, 109.

dad que se nos ofrezca y que nos induzca a hacer el informe.

Pero este principio consagra, ciertamente, la noción cartesiana de que «nada está más cerca de la mente que ella misma», e implica toda una epistemología y metafísica, que sería específicamente dualista [5]. Por eso, no es extraño, una vez que hemos concentrado esta concepción en la idea de «propiedad fenoménica», que «la explicación materialista... no pueda encontrar un lugar para el hecho de que nuestra aprehensión imperfecta se haga mediante una propiedad fenoménica».

Sin embargo, debemos preguntarnos si hay alguna intuición prefilosófica que se contenga en (P) y que se pueda separar de la imagen cartesiana. ¿Cuál es, exactamente, la diferencia entre describir erróneamente algo como una estrella y describir erróneamente algo como un dolor? ¿Por qué lo primero parece una posibilidad evidente y lo segundo parece inimaginable? Quizá la respuesta sea parecida a ésta: Esperamos que la estrella tenga el mismo aspecto aun después de que nos demos cuenta de que es una bola de fuego muy lejana y no un agujero próximo, pero el dolor debería provocar una sensación diferente en el momento en que nos diéramos cuenta de que es una fibra-C estimulada, pues el dolor *es* una sensación, mientras que la estrella *no* es una apariencia visual. Sin embargo, si presentamos esta respuesta, seguimos todavía con la idea de la «sensación» y con el misterio de si los Antípodas tienen alguna sensación. ¿Cuál es la diferencia, debemos preguntarnos, en-

---

[5] George Pitcher ha realizado una descripción de la conducta lingüística que adoptamos al informar de dolores sin utilizar esta premisa. Pitcher estima que los dolores son informes del tejido periférico lastimado, mientras que los Antípodas piensan que son informes de los estados del sistema nervioso central. En su opinión, es un error pensar que el concepto de dolor de sentido común es el concepto de un particular mental. Yo diría que *es* el concepto de un particular mental, pero que sus análisis de la categoría epistemológica de los dolores se aplican, *mutatis mutandis,* con la misma validez siempre que se insiste en esta cuestión. Véase Pitcher, «Pain Perception», *Philosophical Review,* 79 (1970), 368-393. La estrategia general de Pitcher es una defensa del realismo directo, que aparece también en su *A Theory of Perception* (Princeton, 1971) y en D. M. Armstrong, *Perception and the Physical World* (Londres y Nueva York. 1961) y *A Materialist Theory of the Mind* (Londres y Nueva York, 1968). Esta estrategia me parece esencialmente correcta, y capaz de demostrar el carácter opcional de la concepción del particular mental. Pero tengo mis dudas sobre la postura metafilosófica de Pitcher y Armstrong, que convertirían esta idea en una falsa interpretación filosófica de lo que creemos, más que en una explicación correcta de lo que creemos (pero no es necesario que sigamos creyendo).

tre un dolor y reaccionar simplemente a una fibra-C estimulada con el vocablo «dolor», conducta de evitación, y cosas semejantes? En este punto nos inclinamos a decir: absolutamente ninguna diferencia procedente del exterior, y toda la diferencia procedente del interior. La dificultad estará en que nunca habrá una forma en que podamos explicar esta diferencia a los Antípodas. Los Antípodas materialistas creen que no tenemos sensaciones, porque no creen que haya eso que se llama «sensación». Los Antípodas epifenomenalistas creen que puede haberlas, pero no logran imaginarse los motivos por los que podemos armar tanto alboroto para saber si las tenemos. Los filósofos terrestres que piensan que los Antípodas tienen sensaciones pero no lo saben, han llegado a la etapa terminal del filosofar mencionada por Wittgenstein: lo que les apetece es únicamente emitir un sonido no articulado. Ni siquiera pueden decir a los Antípodas que «es diferente para nosotros en el interior», pues los Antípodas no entienden la noción de «espacio interior»; creen que «dentro» significa «dentro del cráneo». *Ahí,* señalan acertadamente, *no es* diferente. Los filósofos terrestres que piensan que los Antípodas no tienen sensaciones están en mejor situación únicamente porque consideran que va contra su dignidad discutir con seres estúpidos sobre si tienen mente.

Parece que siguiendo la objeción presentada por Brandt y Campbell no llegamos a ninguna parte. Probemos por otro camino. En la concepción materialista, toda apariencia de algo va a ser, en realidad, un estado del cerebro. Por eso, parece, el materialista tendrá que decir que el «tosco» duplicado de un estado del cerebro (cómo se encuentran las fibras-C estimuladas) deberá ser otro estado cerebral. Pero, podríamos decir entonces, que ese otro estado cerebral sea el referente de «dolor» en vez de las fibras-C estimuladas. Cada vez que el materialista diga «pero ésa es precisamente nuestra descripción de un estado cerebral», su oponente le contestará: «Muy bien, vamos a hablar del estado cerebral que es el 'acto de aprehensión imperfecta' del primer estado cerebral» [6]. De esta manera el materialista se ve obligado a retroceder cada vez más —apareciendo lo mental siempre que aparece el error. Es como si la Esencia de Vidrio del hombre, el Espejo de la Naturaleza, sólo se hiciera visible a sí misma cuando estuviera ligeramente nublada. Un sistema neural no puede tener nubes, pero una mente sí. Por eso, las mentes, concluimos nosotros, no pueden ser sistemas neurales.

---

[6] Debo esta forma de presentar la postura de Brandt-Campbell a Thomas Nagel.

Veamos ahora cómo considerarían los Antípodas los «actos de aprehensión imperfecta». Los concebirían no como porciones empañadas del Espejo de la Naturaleza, sino como resultado del aprendizaje de un lenguaje de segunda categoría. Toda la idea de las entidades cognoscibles incorregiblemente, en oposición a ser incorregible respecto a cómo parece que son las entidades —la idea de los «pareceres» como si fueran en sí mismos una especie de entidad— la consideran como una forma extraña y deplorable de hablar. El vocabulario terrestre de los «actos de aprehensión», «estados cognitivos», «sensaciones», etc., lo consideran como una dirección poco afortunada seguida por el lenguaje. Creen que no hay forma de escapar, como no sea proponiendo que enseñemos a algunos de nuestros niños a hablar en el lenguaje de los Antípodas y luego veamos si sus resultados son tan buenos como los de un grupo de control. Los Antípodas materialistas, en otras palabras, entienden nuestra idea de «la mente y la materia» como el reflejo de un desarrollo lingüístico poco afortunado. Los Antípodas epifenomenalistas están desconcertados ante la pregunta: «¿Cuál es la entrada neural al centro del habla terrestre que produce informes de dolor así como informes de la fibra-C?» Los filósofos terrestres que piensan que los Antípodas tienen sentimientos creen que el lenguaje de los Antípodas es «inadecuado a la realidad». Los filósofos terrestres que opinan que los Antípodas no tienen sensaciones se basan en una teoría del desarrollo lingüístico según la cual las cosas que se nombran primero son las cosas «que mejor conocemos» —las sensaciones primarias— por lo que la ausencia de un nombre para la sensación implica la ausencia de sensación.

Para precisar un poco más la cuestión, podríamos prescindir de los Antípodas epifenomenalistas y de los terrestres escépticos. El problema de los primeros sobre la neurología de los informes de dolor parece insoluble; para que puedan continuar su tarea caritativa de atribuir a los terrestres estados que son desconocidos para los Antípodas, deberán aceptar todo un sistema dualista, irrefutable con nuevas invesigaciones empíricas, para explicar nuestra conducta lingüística. En el caso de la afirmación de los terrestres escépticos, según la cual los Antípodas no tienen sensaciones primarias, está basada totalmente en la afirmación a priori de que no se puede tener una sensación primaria y no tener una palabra para la misma. Ninguna de las dos posiciones intelectuales —la excesiva benevolencia del Antípoda epifenomenalista y la desconfianza pueblerina del terrestre escéptico— resulta atractiva. Nos quedamos, por una parte, con el Antípoda materialista que dice: «Se creen que tienen

sensaciones pero no las tienen» y, por la otra, con los filósofos terrestres que dicen: «Tienen sensaciones pero no lo saben.» ¿Hay alguna forma de salir de este punto muerto, dado que todo resultado empírico (conexión de cerebros, etc.) parece repercutir igualmente en ambos lados? ¿Existen métodos filosóficos poderosos que penetren en el problema y o lo solucionen o presenten algún compromiso acertado?

3. Incorregibilidad y sensaciones primarias

Un método filosófico que no servirá absolutamente para nada es el «análisis de los significados». Todos entienden perfectamente lo que quieren decir todos los demás. El problema es que un bando piensa que hay demasiados significados y el otro que son demasiado pocos. En este sentido, la analogía más semejante que se puede encontrar es el conflicto entre teístas inspirados y ateos no inspirados. Un teísta inspirado, vamos a suponer, es aquel que «sólo sabe» que hay seres sobrenaturales que desempeñan ciertas funciones explicativas al dar cuenta de los fenómenos naturales. (No deben confundirse con los teólogos naturales —que presentan lo sobrenatural como la mejor explicación de estos fenómenos). Los teístas inspirados han heredado una imagen del universo en que éste aparece dividido en dos grandes esferas ontológicas —la sobrenatural y la natural— y un lenguaje. Su forma de hablar sobre las cosas está indisolublemente unida —o al menos a ellos les parece que está indisolublemente unida— con las referencias a lo divino. La idea de lo sobrenatural no les parece una «teoría», lo mismo que la idea de lo mental no nos parece a nosotros una «teoría». Cuando se encuentran con ateos los consideran como personas que no saben lo que pasa, aunque admiten que los ateos parecen capaces de predecir y controlar los fenómenos naturales con mucha perfección. («Gracias a Dios», dicen, «no somos como esos teólogos naturales, pues, si no, también nosotros podríamos perder contacto con lo real».) Los ateos consideran que estos teístas tienen demasiadas palabras en su lenguaje y demasiados significados de que ocuparse. Los ateos animosos explican a los teístas inspirados que «todo lo que hay *realmente* es...», y los teístas responden que habría que tener en cuenta que hay más cosas en el cielo y en la tierra. ...Y así les va. Los filósofos de cada bando pueden analizar significados hasta quedar sin fuerzas, pero todos estos análisis son o «direccionales» y «reductivos» (por ejemplo, análisis «no-cognitivos»

del discurso religioso, que son análogos a las teorías «expresivas» de los informes de dolor) o solamente describen «formas de vida» alternativas, que vienen a culminar en algo cuya utilidad nunca será mayor que la de anunciar: «Se ha terminado este juego lingüístico.» El juego de los teístas es esencial para su auto-imagen, lo mismo que la imagen de la Esencia de Vidrio del hombre es esencial para la del intelectual occidental, pero ni unos ni otros tienen un contexto más amplio en que puedan evaluar esta imagen. En definitiva, ¿de dónde podría venir este contexto?

Quizá de la filosofía. Cuando no sirven el experimento ni el «análisis del significado», los filósofos se han orientado tradicionalmente a la construcción de sistemas —inventar un nuevo contexto sobre la marcha, por así decirlo. La estrategia más habitual es llegar a un compromiso que haga posible que tanto los que están a favor del principio de Occam, de no multiplicar los seres sin necesidad (por ejemplo, los materialistas, los ateos) como los que se aferran a lo que «saben sin más» sean considerados indulgentemente como si hubieran conseguido «perspectivas alternas» de alguna realidad más amplia que la filosofía acaba de bosquejar. Por eso, algunos filósofos más benévolos han superado la «guerra entre ciencia y teología» y han considerado que Buenaventura y Bohr poseen «formas de conciencia» diferentes, no competitivas. A la pregunta «conciencia ¿de *qué?*» se responde algo parecido a «el mundo» o «la cosa-en-sí-misma» o «la diversidad sensible» o las «estimulaciones». No importa cuál de las respuestas se presente, pues todas ellas son términos destinados a dar nombre a entidades que no tienen ningún rasgo interesante, excepto el de la tranquilidad neutral. Una táctica análoga, entre los filósofos de la mente de tendencia más dura, es el monismo neutral, en que lo mental y lo físico se presentan como dos «aspectos» de alguna realidad subyacente que no necesita mayor descripción. Algunas veces se nos dice que esta realidad es intuida (Bergson) o es idéntica a la materia prima de la sensación (Russell, Ayer), pero en otras ocasiones simplemente se postula como único medio de evitar el escepticismo epistemológico (James, Dewey). En ningún caso se nos dice nada sobre ella, excepto que «sabemos cómo es» o que lo exige la razón (es decir, la necesidad de evitar dilemas filosóficos). A los monistas les gusta señalar que la filosofía ha descubierto, o debería buscar, un substrato subyacente, de la misma manera en que el científico ha descubierto moléculas, y así sucesivamente. Pero, de hecho, no se comprueba que la «sustancia neutra» que no es ni mental ni física tenga capacidades ni propiedades, sino que simplemente se postulan y luego se olvidan

(o, lo que viene a ser lo mismo, reciben el papel de dato inefable)[7]. Esta táctica no puede ayudar a solucionar la cuestión planteada por los filósofos terrestres de línea dura en relación con los Antípodas: ¿Tienen sensaciones primarias o no?

El problema de los Antípodas se puede resumir de la siguiente manera:

1. Un rasgo esencial de las sensaciones primarias es que sean incorregiblemente cognoscibles

Esto, junto con

2. No hay nada en lo que los Antípodas se crean incorregibles

parece llevarnos a

3. Los Antípodas no tienen sensaciones primarias

o a

4. Los Antípodas no tienen conocimiento de su propio conocimiento incorregible.

El problema de (3) es que los Antípodas tienen prácticamente la misma conducta, fisiológica y cultural que nosotros. Además, podemos enseñar a los niños Antípodas a informar de las sensaciones primarias, y a considerarse incorregibles al respecto. Estas consideraciones parecen llevarnos a (4). Pero (4) parece absurdo, y, cuando menos, habría que suavizarlo para convertirlo en

[7] Instando a los filósofos a hacer algo más que esto, Cornelius Kampe ha insinuado que la teoría de la identidad mente-cuerpo sólo tendrá sentido si presentamos «un marco de referencia teórico (o una ontología) de tales características que nos ofrezca una conexión entre los dos fenómenos diversos cuya identidad se afirma». El motivo de esta renovación del monismo neutro es su creencia de que ver el sentido de una teoría de la identidad exige que «se abandone la distinción subjetivo-objetivo, así como la condición privilegiada de los informes introspectivos en primera persona». Este cambio, dice él, «afectaría drásticamente a la lógica de nuestro lenguaje». Creo que Kampe tiene razón en que el abandono de la distinción subjetivo-objetivo tendría resultados verdaderamente drásticos, pero se equivoca al pensar que lo tendría el abandono del acceso privilegiado. Como creo que ha demostrado Sellars, y como estoy intentando demostrar aquí, la distinción subjetivo-objetivo (la idea de «pareceres») puede arreglárselas perfectamente sin las ideas de «mente», «propiedad fenoménica», etc. (cf. Kampe, «Mind-Body Identity: A Question of Intelligibility», *Philosophical Studies,* 25 [1974], 63-67).

4'. Los Antípodas no tienen conocimiento de su propia capacidad de conocimiento incorregible

que resulta un poco extraño, pero al menos tiene algunos paralelos. (Compárese con «Juan XXIII tuvo que ser convencido con argumentos de su propia infalibilidad cuando accedió al papado».) Sin embargo, si insistimos en (4'), la posibilidad de enseñar a los niños antípodas parece que nos deja colgados entre

5. Se puede enseñar a los Antípodas a reconocer sus propias sensaciones primarias

y

5'. Garcias a la presencia de concomitantes neurales de las sensaciones primarias, se puede enseñar a los Antípodas a simular informes de sensaciones primarias sin llegar a tenerlas en la realidad.

Podríamos tener la esperanza de resolver este nuevo dilema encontrando un Antípoda bilingüe. Pero el bilingüe no tiene conocimiento «interior» sobre los significados de las expresiones de la lengua extranjera; tiene el mismo tipo de teoría que el que realiza un léxico. Imaginemos un Antípoda adulto que llega a hablar uno de nuestros idiomas. Dice «Tengo dolor» o el equivalente antípoda «Se están activando mis fibras-C», según de qué esté hablando. Si un interlocutor terrestre le dice que no tiene realmente dolores señala que la observación es una alocución desviada, y exige acceso privilegiado. Cuando los interlocutores antípodas le hacen ver que no se están activando realmente sus fibras-C, dice algo parecido a «Tiene gracia; es indudable que parecen estarlo. Por eso dije a los terrestres que tenía dolor», o quizá algo parecido a «Tiene gracia; es indudable que tengo lo que los terrestres llaman 'dolor', y eso no ocurre nunca, a no ser que se estén activando mis fibras-C». Es difícil imaginar que una u otra fórmula mereciera claramente nuestra preferencia, y todavía más difícil pensar que los filósofos pudieran conseguir algo con una preferencia, si la tuvieran. Una vez más, parece que estamos ante la pregunta retórica «Pero, ¿qué impresión produce?» —a lo que el Antípoda bilingüe responde: «De dolor.» Cuando se le pregunta si no produce una sensación semejante a la de las fibras-C, explica que en su lenguaje no hay equivalente de «sensación» y por eso no se le ocurriría decir que

tenía la sensación de que se activaban sus fibras-C, aunque tiene conciencia de ello siempre que ocurre ese hecho.

Si esto parece paradójico, hay que suponer que se debe a que creemos que «conciencia no inferencial» y «sensación» son casi totalmente sinónimos. Pero no sirve de mucho señalar esto. Si los consideramos como sinónimos, entonces es lógico que el Antípoda tenga el concepto de un estado llamado «sentir», aunque, sin embargo, no tenga el concepto de «sensaciones» en cuanto objetos intencionales del conocimiento. El Antípoda tiene el verbo pero no el nombre, por así decirlo. Un Antípoda acomodaticio puede decir que su lenguaje puede expresar la idea de «un estado tal que nadie pueda confundirse al pensar que está en él» —a saber, el estado de que a uno le parezca que...— y, sin embargo, tener dudas sobre si dichos estados son lo mismo que los dolores y las otras sensaciones primarias en que tan interesados están los terrestres. Por una parte, parece que es de todo lo que *pueden* hablar, pues recuerda que ha aprendido a decir «dolor» cuando, y sólo cuando, parece que se activan sus fibras-C. Por otra parte, los terrestres insisten en que hay una diferencia entre estar en un estado tal que a uno le parezca que está... y tener una sensación primaria. El primer estado es una posición epistémica hacia algo sobre lo que es posible dudar. El último estado coloca automáticamente en una posición epistémica hacia algo sobre lo que es imposible dudar.

Así, pues, el dilema parece quedar reducido a esto: Tenemos que afirmar o negar

6. cualquier informe de cómo algo parece, es un informe de una sensación primaria.

La única razón para afirmarlo parece ser que es corolario de la proposición recíproca de (1), es decir:

7. Es esencial a todo lo que sea incorregiblemente cognoscible que sea una sensación primaria.

Pero (7) no es más que una forma del principio invocado por la anterior objeción de Brandt-Campbell, a saber:

(P) Siempre que hagamos un informe incorregible sobre un estado de nosotros mismos, tiene que haber una propiedad que se nos ofrezca y que nos induzca a hacer el informe

y en este principio todo depende de la idea de «que se nos ofrece» —idea que nos recuerda inmdiatamente las metáforas del «Ojo de

la Mente», «presencia a la conciencia», y semejantes, que a su vez proceden de la imagen inicial del Espejo de la Naturaleza— del conocimiento en cuanto conjunto de representaciones inmateriales. Si adoptamos este principio, entonces, por extraño que parezca, ya no podemos ser escépticos: los Antípodas tienen automáticamente sensaciones primarias. Debemos preferir (5) a (5'). Como no ponemos en duda que a algunos Antípodas les parezca que tienen calambres en el estómago o que se activen sus fibras-C, y como reconocemos la incorregibilidad de tales informes, tenemos que reconocer que tienen algunas sensaciones primarias que son la «base» de sus afirmaciones-pareceres y de las que se les podría enseñar a informar aprendiendo un vocabulario apropiado. Pero esto significa, paradójicamente, que cierta especie de conductismo se deriva del mismo principio que encarna la imagen cartesiana del Ojo de la Mente —la misma imagen que ha recibido muchas veces la acusación de llevar al «velo de ideas» y al solipsismo. Sólo podríamos ser escépticos y afirmar (5') —que lo más que podrían hacer los Antípodas sería simular— afirmando que cuando los Antípodas hacen afirmaciones de pareceres no quieren decir realmente lo que nosotros queremos dar a entender con ellas, y que el desvío, en el idioma de los Antípodas, de la expresión «Puedes estar equivocado al decir que te parece que se están activando tus fibras-C» no basta para demostrar que los Antípodas tengan ningún conocimiento incorregible. Es decir, deberíamos reinterpretar la conducta que consideramos que se había producido inicialmente, y basar nuestro escepticismo sobre sus sensaciones primarias en un escepticismo más general sobre su posesión de conocimiento (o de algunas clases de conocimiento). Pero no es fácil ver cómo podríamos hacer plausible el escepticismo a este respecto, como no sea basándonos en alguna convicción anterior de que no tenían mente —convicción que excluiría, a priori, las sensaciones primarias. Por eso, en este caso el escepticismo habrá de ser gratuito y pirroniano[8]. Por otra parte, si negamos (6) —si separamos el parecer del

[8] Conviene distinguir entre el escepticismo «simple» o pirroniano, y la forma de escepticismo específicamente «cartesiana» que invoca el «velo de ideas» como justificación de una actitud escéptica. El escepticismo «pirroniano», en el sentido que yo daré al término, dice simplemente: «No podemos estar nunca seguros; así que, ¿cómo podemos llegar a saber?» Por el contrario, el escepticismo del «velo de ideas» dice algo más específico, a saber: «Dado que nunca tendremos certeza de nada, excepto de los contenidos de nuestras propias mentes, ¿cómo vamos a llegar nunca a justificar una inferencia de una creencia sobre cualquier otra cosa?» Véase una exposición del entrelazamiento de estas

tener estados mentales y abandonamos las imágenes cartesianas— tenemos que enfrentarnos con la posibilidad de que nosotros mismos no tuviéramos sensaciones, ni estados mentales, ni mentes, ni Esencia de Vidrio. Esta paradoja parece tan fuerte que nos haría volver sin más a (P) y al Espejo de la Naturaleza.

Así pues, el problema se reduce a una elección entre tres posibilidades inquietantes. Tenemos que compartir nuestra Esencia de Vidrio con todo ser que parezca hablar un lenguaje que contenga afirmaciones-de-pareceres, o volvernos escépticos pirronianos, o enfrentarnos con la posibilidad de que nunca tengamos esta esencia. Si admitimos (7) —la premisa que que ser una sensación primaria sea esencial a ser un objeto de conocimiento incorregible— entonces hemos de admitir o (a) que el lenguaje de los Antípodas, precisamente por contener algunos informes incorregibles, versa sobre sensaciones primarias, o (b) que nunca llegaremos a saber si los Antípodas hablan una lengua precisamente porque nunca llegaremos a saber si tienen sensaciones primarias, o (c) que toda la cuestión de las sensaciones primarias es un falso problema, pues el ejemplo de los Antípodas demuestra que nosotros no hemos tenido nunca sensaciones primarias.

Estas tres posibilidades se corresponden a grandes rasgos con las tres posturas clásicas en la filosofía de la mente —conductismos, escepticismo sobre las otras mentes, y materialismo. En vez de aceptar cualquiera de ellas, yo sugiero que rechacemos (7), y con ella (P). Es decir, que abandonemos la idea de que poseemos conocimientos incorregibles en virtud de una relación especial con un tipo especial de objeto llamado «objetos mentales». Esta sugerencia es un corolario del ataque de Sellars al Mito de lo Dado. Presentaré este ataque más detalladamente en el capítulo cuatro, pero en este momento quiero señalar simplemente que este mito es la idea de que relaciones epistémicas como «conocimiento directo» o «conocimiento incorregible» o «conocimiento cierto» deben entenderse según un modelo causal, para-mecánico, como una relación especial entre ciertos objetos y la mente humana que hace posible que el conocimiento se produzca con más facilidad, naturalidad o rapidez Si pensamos en el conocimiento incorregible simplemente como una cuestión de práctica social —de la ausencia, en la conversación normal, de réplica a una determinada pretensión de conocimiento— nunca parecerá plausible un principio como (7) o (P).

En las dos secciones anteriores me he referido al «objeto mental»

---

dos formas de escepticismo en Richard Popkin, *The History of Scepticism from Erasmus to Descartes* (Nueva York, 1964).

como si fuera sinónimo de «objeto cognoscible de forma incorregible» y, por tanto, como si tener mente fuera lo mismo que tener conocimiento incorregible. He dejado de lado la inmaterialidad y la capacidad de abstraer, que fueron examinadas en el capítulo primero, y la intencionalidad, de la que me ocuparé en el capítulo cuarto. Mi excusa para pretender que la mente no es más que un conjunto de sensaciones primarias incorregiblemente introspectibles, y que su esencia es esta categoría epistémica especial, es que esa misma pretensión está generalizada en toda el área llamada «filosofía de la mente». Esta área de la filosofía ha visto la luz en los treinta años transcurridos a partir de la publicación de *The Concept of Mind,* de Ryle. El resultado de ese libro fue hacer que las cuestiones de las mentes y cuerpos giraran casi totalmente en torno a los casos que ofrecían resistencia al intento conductista lógico del propio Ryle por acabar con el dualismo cartesiano —a saber, las sensaciones primarias. El estudio de las sensaciones realizado por Wittgenstein en *Investigaciones filosóficas* parecía ofrecer el mismo tipo de intento de supresión. Por eso, muchos filósofos han dado por hecho que «el problema mente-cuerpo» era la cuestión de si las sensaciones primarias podían considerarse como disposiciones a comportarse. Y así, ha dado la impresión de que las únicas posibilidades eran las que acabo de mencionar: (a) admitir que Ryle y Wittgenstein tenían razón, y que no hay objetos mentales, (b) decir que estaban equivocados, y que, por lo tanto, el dualismo cartesiano se mantiene intacto, siendo una consecuencia natural el escepticismo sobre las otras mentes, y (c) alguna forma de teoría de la identidad mente-cuerpo, según la cual Ryle y Wittgenstein estaban equivocados, pero eso no significa que haya que reivindicar a Descartes.

El resultado de plantear los problemas de esta manera es centrar la atención en los dolores, y prestar menos atención al aspecto de la mente que es, o debería ser, de más importancia para la epistemología —las creencias e intenciones. (En los últimos años se ha restablecido en cierta medida el equilibrio gracias a los filósofos de la mente que tratan de tender puentes hacia la psicología empírica. Hablaremos de su trabajo en el capítulo quinto.) Pero sigue siendo verdad que se cree que «el problema mente-cuerpo» es en primer lugar un problema sobre dolores, y el aspecto distintivo de los dolores es precisamente el mencionado por Kripke —que parece que no hay, en relación con nuestro conocimiento de los mismos, algo que se pueda considerar una distinción apariencia-realidad. De hecho, como he tratado de hacer ver en el capítulo primero, éste es solamente uno entre los varios «problemas mente-cuerpo», cada

uno de los cuales ha contribuido a crear la noción confusa de que hay algo especialmente misterioso en el hombre que le hace capaz de conocer, o de ciertas formas especiales de conocer.

No obstante, en lo que queda de capítulo voy a tratar de justificar mi propuesta de abandonar (P) y, por tanto, no ser ni dualistas, escépticos o conductistas, ni «teóricos de la identidad». No sé cómo argumentar directamente contra (P), pues la afirmación de que el conocimiento incorregible consiste en el ofrecimiento de una propiedad fenoménica no tanto una afirmación cuanto un resumen de toda una teoría —todo un conjunto de términos y suposiciones que se centran en torno a la imagen de la mente en cuanto que refleja la naturaleza, y que contribuyen a dar un sentido a la afirmación cartesiana de que la mente se «da» a sí misma naturalmente. Es esta imagen misma la que hay que dejar de lado, si queremos penetrar en la idea del siglo XVII de que podemos comprender y mejorar nuestro conocimiento comprendiendo el funcionamiento de nuestra mente. Espero demostrar la diferencia existente entre dejarla de lado y adoptar cualquiera de las posturas que presuponen esta imagen. Por tanto, el resto del capítulo estará dedicado al conductismo, al escepticismo y a la teoría de la identidad mente-cuerpo, en un intento de diferenciar mi postura de cada una de las mencionadas. En la conclusión del capítulo —«Materialismo sin identidad»— trato de decir algo más positivo, pero este intento debe ponerse en relación con el estudio de los otros problemas «mente-cuerpo», pues es la única forma de que parezca plausible.

4. El conductismo

El conductismo es la doctrina que afirma que hablar de «estados internos» no es más que una forma abreviada, y quizá engañosa, de hablar de disposiciones a comportarse de determinadas maneras. En su forma ryleana o «lógica» —de la que me ocuparé a continuación— su doctrina central es que hay una conexión necesaria entre la verdad de un informe de una cierta sensación primaria y una disposición a tal o cual conducta. Un motivo para mantener este punto de vista es la desconfianza en lo que Ryle llamaba «fantasmas en las máquinas», la imagen cartesiana de las personas, y otro es el deseo de impedir que el escéptico ante otras mentes plantee la cuestión de si la persona que se retuerce en el suelo tiene sensaciones del mismo tipo que tendría el propio escéptico cuando se revolcara por el suelo. En la concepción conductista lógica, los informes de estas

sensaciones no deben considerarse como si hicieran referencia a entidades no físicas, y quizá a ninguna entidad en absoluto, a no ser al retorcerse o a la disposición a retorcerse.

Esta doctrina ha sido criticada por la razón de que parece que no hay forma de rellenar una descripción de la disposición necesaria y comportarse sin presentar listas infinitamente largas de movimientos y ruidos posibles. También se la ha atacado diciendo que cualquiera que sea la «necesidad» que haya en el área no es cuestión de «significado» sino sencillamente una expresión del hecho de que solemos explicar cierta conducta haciendo referencia a ciertos estados internos —por lo que la necesidad no es más «lingüística» o «conceptual» que la que conecta el color rojo de la estufa con el fuego que hay dentro. Finalmente, ha sido atacada de ser la clase de paradoja filosófica que sólo ocurriría en una mente obsesionada con el dogma instrumentalista o verificacionista —ansiosa por reducir todo lo no observable a observable para evitar cualquier riesgo de creer en algo no real.

Todas estas críticas están, en mi opinión, perfectamente justificadas. Las formulaciones clásicas del conductismo lógico presuponen ciertamente las distinciones entre observación y teoría y entre lenguaje y hecho, que los filósofos, como demostraré en el capítulo cuarto, harían bien en abandonar. Pero sigue en pie la sensación de que el conductista tiene algo de razón. Un punto que tiene a su favor es que parece absurdo pensar que algún día, después de años de fructíferas conversaciones con los Antípodas, tengamos razones para decir: «No hay sensaciones primarias, y, por lo tanto, tampoco hay mentes; por lo tanto, no hay lenguaje, y en definitiva no hay personas.» La idea de que podamos vernos obligados a decir que no tenían sensaciones primarias nos obliga a preguntarnos si podemos llegar a imaginarnos de qué clase de obligación se trataría. Nos hace también darnos cuenta de que aun cuando nos viéramos de alguna manera obligados a ello, es casi seguro que no extraeríamos las inferencias sugeridas. Por el contrario, podríamos comenzar a compartir el desconcierto de los Antípodas sobre las razones que nos habían llevado a preocuparnos tanto por esta cuestión. Comenzaríamos a apreciar la actitud burlona de los Antípodas ante toda la cuestión —la misma actitud con que los polinesios recibían la preocupación de los misioneros por la pregunta «¿Son éstos los descendientes de Sem o de Cam»? El punto fuerte del conductista es que cuanto más se intenta contestarle más absurdas parecen volverse las preguntas del filósofo de línea dura: «Mente: ¿sí o no?». «Sensaciones primarias: ¿sí o no?»

Pero este punto a su favor comienza a perder valor en el momento en que se aplica como tesis sobre las «conexiones necesarias establecidas por el 'análisis de los significados'». La intuición de Ryle se vio frustrada por la epistemología positivista que heredó. En vez de decir que el conocimiento incorregible dependía únicamente de cuáles eran las prácticas de justificación adoptadas por los semejantes (la postura que, en el capítulo cuarto, designaré con el nombre de «conductismo epistemológico»), se vio arrastrado a decir que un cierto tipo de conducta formaba una condición necesaria y suficiente para la adscripción de sensaciones primarias, y que esto constituía un hecho relacionado con «nuestro lenguaje». Entonces tuvo que hacer frente a un problema muy difícil. El hecho de que nuestro lenguaje permitiera hacer inferir la presencia de tales sensaciones hacía difícil, sin volver a caer en el materialismo, negar que hubiera realmente entidades fantasmales de qué informar. De esta manera entraban en conflicto los dos motivos que estaban tras el conductismo lógico, dado que el deseo de encontrar una barrera «lógica» al escepticismo sobre las otras mentes parece llevar otra vez hacia el dualismo. Efectivamente, si aceptamos la idea de que una determinada práctica lingüística, una obra de conducta determinada, es suficiente (a pesar de los escépticos) para demostrar la necesidad de sensaciones primarias en el interior, cualquiera que sea el sentido en que existen las sensaciones primarias, entonces parece necesario afirmar que nuestra experiencia conversacional con los Antípodas implica que tienen sensaciones primarias en el sentido, cualquiera que sea, en que *nosotros* tenemos sensaciones primarias. Es decir, parece necesario adoptar el siguiente punto de vista:

(P') La capacidad de hablar un lenguaje que incluya afirmaciones-de-pareceres incorregibles supone la presencia de sensaciones primarias en los hablantes de esa lengua, en el sentido, cualquiera que sea, en que las sensaciones primarias están presentes en *nosotros*.

Es difícil ridiculizar la idea de que podemos descubrir la verdad de esta afirmación haciendo lo que se llama «analizar significados» [9]. Parece fácil decir (con el escéptico) que podríamos tener la capacidad sin tener las sensaciones. Pero es difícil, como han dejado claro Wittgenstein y Bouwsma, hacer una narración coherente sobre

[9] Véase una muestra de este ridículo en Hilary Putnam, «Brains and Behavior» en *Mind, Language and Reality*, vol. II (Cambridge, 1975).

lo que hemos imaginado. A pesar de esto (P') tiene cierta plausibilidad. La razón de que sea plausible es que, una vez más, es un corolario de:

> (P) Siempre que hacemos un informe incorregible sobre un estado de nosotros mismos, tiene que haber una propiedad que se nos ofrezca y que nos induzca a hacer este informe

—principio vital para la imagen del Espejo de la Naturaleza. Es la imagen según la cual la «apariencia» no es simplemente una creencia equivocada sino la creencia equivocada engendrada por un mecanismo particular (algo engañoso que se pone ante el Ojo de la Mente) que hace que parezca tan necesaria la conexión entre conducta y sensación primaria. Es la imagen según la cual están implicadas tres cosas cuando alguien se equivoca (o, por extensión, acierta): la persona, el objeto de que está hablando, y la representación interna de ese objeto.

Ryle pensaba que había evitado esta imagen, pero se demuestra que no logró hacerlo si observamos su intento por demostrar, paradójica e inútilmente, que no había informes incorregibles. Ryle tenía miedo de que, si existían estos informes, tendría que ser verdad algo parecido a (P) para poder explicar su existencia. Pensaba que si hubiera una capacidad de hacer informes no inferenciales incorregibles sobre los estados interiores, esto demostraría que alguien que no sabía nada de la conducta podría saberlo todo sobre los estados interiores, y por tanto que, en definitiva, Descartes tenía razón. Criticaba, y con razón, la explicación cartesiana habitual de la introspección como un aspecto de la «para-óptica», pero no disponía de ninguna otra explicación y, por tanto, se vio obligado a adoptar la postura insostenible de tener que negar totalmente el fenómeno del acceso privilegiado. Dedicó el capítulo menos convincente de *The Concept of Mind* (el que èstudia el «auto-conocimiento») a la afirmación paradójica de que «las clases de cosas que puedo averiguar sobre otras personas, y los métodos de averiguación son muy similares»[10]. El resultado fue que muchos filósofos que estaban de acuerdo en que Ryle había demostrado que las creencias y deseos no eran estados interiores admitían también que había dejado intactas las sensaciones primarias, y, por tanto, que quedaba por hacer la opción entre dualismo y materialismo[11].

---

[10] Gilbert Ryle, *The Concept of Mind* (Nueva York, 1965), pág. 155. La referencia a la «para-óptica» está en la pág. 159.

[11] Sellars, en «Empiricism and the Philosophy of Mind» (en *Science, Per-*

Podríamos formular con otras palabras el error de Ryle, diciendo que creía que si se podía demostrar una «conexión necesaria entre adscripciones de disposiciones conductuales y adscripciones de estados internos, se habría demostrado que no había realmente estados interiores. Pero se puede evitar esta falsa conclusión instrumentalista, como se puede evitar (P), sin abandonar el criterio antiescéptico de que la conducta de los Antípodas es prueba suficiente para justificar que les atribuyamos tanta o tan poca vida interior como la que poseamos nosotros mismos. La inferencia metafísica que el conductista se ve tentado de hacer —la inferencia de que dentro no hay una Esencia de Vidrio en el interior— es, considerada aisladamente, tan implausible como cualquier otra afirmación instrumentalista. (Compárese: «No hay positrones; sólo hay disposiciones por parte de los electrones para...» «No hay electrones; sólo hay disposiciones por parte de los objetos macroscópicos para...» «No hay objetos físicos; sólo hay disposiciones por parte de los contenidos sensoriales para...»). Despojada de sus pretensiones de rigor, la postura conductista se reduce a recordarnos que la noción de sensación primaria sólo tiene función en el contexto de una imagen que conecte ciertas clases de conducta (informes introspectivos) con otras (informes de objetos físicos) en términos de una determinada imagen de cómo son los seres humanos (no sólo sus mentes). El conductista mira el papel social de la idea del

ception and Reality [Londres y Nueva York, 1963]) dio el primer paso más allá de Ryle. Sellars demostró que, aunque el hecho de que la conducta es prueba de las sensaciones primarias esté «introducido en la lógica» de los conceptos de las sensaciones primarias, esto no quiere decir que no pueda haber sensaciones primarias, lo mismo que la observación paralela sobre macro-fenómenos y micro-entidades no impone las afirmaciones operacionalistas de que no puede haber micro-entidades. En la obra citada Sellars decía lo que Ryle debería haber dicho en su capítulo sobre el «auto-conocimiento», pero no dijo —a saber, que los informes introspectivos no eran más misteriosos que cualquier otro informe no inferencial, y no necesitaba del Mito de lo Dado (y, por tanto, no necesitaba de la para-óptica) para su explicación. Pero, por desgracia, Sellars no extrajo la conclusión de que, como diría más adelante Armstrong, no había eso que se llama «acceso lógicamente privilegiado», sino únicamente «acceso empíricamente privilegiado» (Armstrong, *A Materialist Theory of the Mind* [Londres, 1968], pág. 108). Este punto, que es fundamental para el punto de vista que estoy presentando en este libro, está sugerido en todo lo que dice Sellars, pero fue preciso el ataque de Quine a la distinción lógico-empírico para que fuera posible evitar la fórmula de las «conexiones lógicamente necesarias» que Ryle había introducido en la filosofía de la mente. Como indico en el capítulo cuarto, Sellars no ha sido nunca capaz de asimilar a Quine con toda su fuerza, y cuando hablaba de «la misma lógica de estos conceptos» lo hacía, por desgracia, en la tradición ryleana.

«dolor», sin tratar de profundizar en ella para llegar a la inefable cualidad fenomenológica que tienen los dolores. El escéptico ha de insistir en que es ésta cualidad —que sólo se conoce por propia experiencia— la que cuenta. La razón por la que el conductista se coloca en la paradójica postura metafísica de negar que haya causas mentales no disposicionales de las disposiciones conductuales la expresa Wittgenstein de la siguiente manera:

> ¿Cómo se presenta el problema filosófico sobre los estados y procesos mentales y sobre el conductismo? —El primer paso es el que pasa totalmente desapercibido. Hablamos de procesos y estados y dejamos sin determinar su naturaleza. Es posible que con el tiempo lleguemos a saber más sobre ellos —pensamos. Pero es eso precisamente lo que nos lleva a una forma concreta de considerar la cuestión, pues tenemos un concepto determinado de qué es lo que significa aprender a conocer mejor un proceso. (Se ha realizado el paso decisivo en el juego de prestidigitación, y ha sido precisamente el que nos parecía totalmente inocente)[12].

Alan Donagan comenta admirablemente este pasaje, diciendo que

> los cartesianos... transformaron los hechos gramaticales que resumíamos en la proposición de que la sensación es no-disposicional y privada en la ficción gramatical de que las sensaciones son estados o procesos en un medio privado y, por tanto, no-material.

Los conductistas, por otra parte,

> bien movidos por la aridez de gran parte de la psicología introspeccionista, por las dificultades filosóficas del cartesianismo o por otras consideraciones, comenzaron negando que existan de ninguna manera en su medio no-material los procesos cartesianos privados.

Wittgenstein aclaró la cuestión, piensa Donagan, admitiendo «que las sensaciones son acompañamientos privados y no-disposicionales de la conducta por la que se expresan naturalmente», pero negándose «a reconocer dichos acompañamientos como procesos que puedan recibir un nombre e investigarse independientemente de

---

[12] Ludwig Wittgenstein, *Philosophical Investigations* (Londres y Nueva York, 1953), pt. I, sec. 308.

las circunstancias que los producen, y de la conducta por la que se expresan naturalmente»[13].

Creo que la concisa formulación que hace Donogan de las dificultades comunes de conductas y escépticos dualistas es correcta, pero que se puede clarificar y hacerle dar un paso más. La idea de un «medio privado, ... no material» es oscura, pues da a entender que no tenemos idea de cómo es eso de tener una Esencia de Vidrio —una comprensión metafísica de cómo es la sustancia no extensa— que sea independiente del criterio epistémico de lo mental. Si dejamos de lado esta idea y hacemos hincapié en las palabras de Wittgenstein «un concepto determinado de lo que significa aprender a conocer mejor un proceso podemos obtener un diagnóstico de lo que Donagan llama 'errores opuestos y complementarios'»[14] del conductismo y cartesianismo, en que se evite hacer referencia a la naturaleza metafísica («no-material», fantasmal) de las sensaciones primarias.

La premisa epistemológica básica que comparten ambas escuelas, y que forma su idea de «conocer mejor», es la doctrina de lo Naturalmente Dado, es decir:

> El conocimiento es o una de esas entidades que están naturalmente preparadas para ser inmediatamente presentes a la conciencia, o una entidad cuya existencia y propiedades están implicadas por las entidades de la primera clase (y que, por tanto, son «reducibles» a las primeras).

Los cartesianos creían que las únicas entidades que estaban naturalmente preparadas para ser inmediatamente presentes a la conciencia eran los estados mentales. Los conductistas, en sus mejores momentos epistemológicos, creían que la única clase de entidades directamente presentes a la conciencia eran estados de los objetos físicos. Los conductistas se vanagloriaban de escapar a las ideas de nuestra Esencia de Vidrio y del Ojo Interior pero seguían fieles a la epistemología cartesiana al conservar la idea de un Ojo de la Mente que obtenía algunas cosas de primera mano. La ciencia, en esta concepción, infiere otras cosas implicadas por las entidades «de abajo», y luego la filosofía vuelve a reducir otra vez estas cosas. Los conductistas abandonaron la idea de «nada es mejor conocido

[13] Las cuatro citas son de Alan Donagan, «Wittgenstein on Sensation», *Wittgenstein: The Philosophical Investigations: A Collection of Critical Essays,* ed. George Pitcher (Nueva York, 1966), pág. 350.

[14] *Ibid.,* pág. 349.

a la mente que ella misma» pero conservaron la idea de que algunas cosas eran directamente cognoscibles por naturaleza y otras no, y el corolario metafísico de que sólo las primeras eran «realmente reales». Esta doctrina —que lo más cognoscible era lo más real— que George Pitcher ha denominado «Principio Platónico» [15], sumada al principio de lo Naturalmente Dado, producía o una reducción idealista o panpsíquica de lo físico a lo mental, o una reducción conductista o materialista en el sentido contrario. La elección entre las dos formas de reducción depende, creo yo, no tanto de las dificultades psicológicas o filosóficas cuanto de la propia idea general de cómo es la sabiduría, y, por lo tanto, de para qué sirve la filosofía. ¿Es para resaltar los aspectos del hombre a los que se tiene acceso por métodos públicos de conversación normal e investigación científica? ¿O, más bien, un sentido personal e inarticulado de «algo mucho más profundamente fundido»? Esta opción tiene poco que ver con la argumentación filosófica o con la imagen del Espejo de la Naturaleza. Pero la imagen —y especialmente la metáfora del Ojo Interior— sirve igual de bien para los objetivos de ambos bandos, lo cual es la razón de que la discusión entre ellos haya sido tan prolongada y tan poco concluyente. Ambos tienen un sentido claro de qué es lo que mejor se conoce, y conocer un proceso significa o conocerlo de *esa* manera, o demostrar que «realmente no es más que» alguna otra cosa que se conoce de esa manera.

Si consideramos la controversia entre los conductistas y los escépticos sobre las otras mentes desde el punto de vista de los Antípodas, lo primero que observamos es que no queda lugar para lo «Naturalmente Dado». No hay ningún lugar para la idea de «conocimiento directo». Este es sencillamente el conocimiento que se tiene sin que el poseedor haya realizado ninguna inferencia consciente. Pero no hay ninguna insinuación de que algunas entidades estén especialmente bien preparadas para ser conocidas de esta manera. Lo que conocemos no inferencialmente depende de qué es aquello con lo que estemos familiarizados. Algunas personas (las que están sentadas delante de las cámaras de niebla) están familiarizadas con las partículas elementales, y hacen informes no inferenciales de las mismas. Otras están familiarizadas con las enfermedades de los árboles, y pueden informar «otro caso de enfermedad del olmo holandés» sin realizar ninguna inferencia. Todos los Antípodas están familiarizados con los estados de sus nervios, y todos los Terrestres con sus sen-

[15] Cfr. Pitcher, *Theory of Perception*, pág. 23, y Platón, *La república*, 478B: «Así, pues, si lo real es el objeto del conocimiento, entonces el objeto de la creencia ha de ser algo que no sea lo real.»

sacione primarias. Los Antípodas no pretenden que haya algo de sospechosamente metafísico o fantasmagórico en las sensaciones primarias —sencillamente no ven para qué sirve hablar de estas cosas en vez de hablar de los nervios. Tampoco sirve, naturalmente, que los terrestres expliquen que aunque (dejando de lado la posibilidad de inferencia inconsciente) sea posible que todo *pueda* ser conocido no inferencialmente, no se deduce que se pueda conocer *incorregiblemente* todo, excepto ciertas realidades adecuadas por naturaleza. Los Antípodas no tienen la idea de *entidades* conocidas incorregiblemente sino únicamente de *informes* (afirmaciones-de-pareceres) que son incorregibles y que pueden versar sobre *cualqueir* clase de entidad. Comprenden que los Terrestres no tengan la primera idea, pero les intriga por qué piensan que la necesitan, aunque se hacen cargo de que, cuando se ignora la neurología, pueden ponerse en circulación muchas ideas extrañas.

5. Escepticismo sobre las otras mentes

Cuando volvemos la atención al escéptico sobre las otras mentes —la persona que insiste en que puede haber estados interiores sin acompañamientos conductuales—, vemos otra vez cómo una intuición perfectamente válida se transforma en una paradoja. La intuición válida es que las sensaciones primarias son particulares tan legítimos como las mesas o los arcángeles o los electrones —habitantes tan legítimos del mundo como aspirantes legítimos a una categoría ontológica. La pregunta «¿Tienen sensaciones primarias los Antípodas?» no es más confusa ni más «metafísica» que la pregunta «¿Tienen la sangre roja?» o «¿Tienen un sentido moral?». Además, nosotros tenemos una forma especial y superior de tener conocimiento de nuestras propias sensaciones primarias —tenemos acceso privilegiado a las entidades privadas.

Esta afirmación válida comienza a torcerse en el momento en que el principio de lo Naturalmente Dado sugiere que, puesto que las sensaciones primarias son realmente tan bien conocidas, tienen que ser entidades de una clase muy especial —procesos en un «medio no-material» privado, quizá. Se estropea, por así decirlo, cuando se considera que la Esencia de Vidrio del hombre es algo más que el hecho de que el conocimiento del hombre sea un Espejo de la Naturaleza, y cuando se plantea la pregunta «¿Qué clase de material, o no-material, especial y maravilloso, podría realizar ese reflejo?». No hay nada de malo en decir que las personas tienen que

ser muy especiales para saber tantísimas cosas más que las bestias del campo; hasta los Antípodas lo dicen. Pero cuando tratamos de pasar de

1. Conocemos nuestras mentes mejor que ninguna otra cosa

a

2. Podríamos saberlo todo sobre nuestras mentes aun cuando no supiéramos nada más

a

3. El saber si algo tiene mente es cuestión de saber si se conoce a sí mismo

no podemos decir nunca por qué no vamos a ser solipsistas. La transición de (1) a (2) resulta natural, aunque no necesaria, en virtud del principio de lo Naturalmente Dado y la metáfora del Ojo Interior. Si pensamos en que el Ojo se vuelve simplemente hacia adentro y detecta una sensación primaria, todo el complejo de instituciones sociales y manifestaciones conductuales que rodean a los informes de tales sensaciones primarias parece irrelevante. Precisamente por ser irrelevante, nos vemos impulsados a pasar de (2) a (3) —lo único que llegaremos a saber sobre nuestros semejantes (si los tenemos), podemos decir ahora con tisteza, es su conducta y su posición social. Nunca sabremos lo que ocurre en su interior, si es que hay realmente algo allí «dentro». El resultado de esto es que dejamos de pensar en nuestros amigos y vecinos en cuanto personas y comenzamos a pensar en ellos como cáscaras que rodean una cosa misteriosa (la Esencia de Vidrio, el medio privado no material) que sólo pueden llegar a describir los filósofos profesionales, pero que nosotros sabemos que está allá, o al menos eso esperamos. La idea de que la introspección es un vislumbre de otro campo ontológico no es (y en esto los wittgensteinianos tienen toda la razón) algo que resulte obvio cuando hacemos de hecho la introspección. Cuando dirigimos hacia dentro el Ojo de la Mente no nos invade ninguna sensación de espiar algo misterioso. La idea de que lo estemos haciendo es producto de las nociones epistemológicas que nos dejaron pasar de (1) a (2) y a (3). Aquí, como en todas partes, la epistemología precede a la metafísica y nos impulsa hacia ella.

Pero la extrañeza artificial producida por un exceso de episte-

mología no debería llevarnos (y en esto se equivocan completamente algunos wittgensteinianos) a pensar que no puede haber nada dentro. Ni tampoco deberíamos pensar que nuestro acceso privilegiado a nuestros propios estados mentales sea un misterio que necesite o una defensa metafísica o una destrucción escéptica. La fuerza de intuición original válida del escéptico puede resaltarse insistiendo en este último punto. Yo lo haré examinando críticamente los argumentos wittgensteinianos de que las entidades mentales tienen una especie de responsabilidad ontológica menor, y que toda la idea de entidades «privadas» y acceso privilegiado a las mismas está mal enfocada. Strawson ha subrayado que Wittgenstein tenía dos clases diferentes de «hostilidad» —una a lo privado y otra a la inmediatez. Esta última, pienso yo, era de una importancia fundamental, pero la primera estaba totalmente descaminada [16].

Consideremos el famoso pasaje en que Wittgenstein dice que una sensación «no es un *algo,* pero tampoco una *nada* (!). La conclusión era únicamente que una nada podría servir igual que un algo sobre el que no se pudiera decir nada» [17]. Esto dirige la atención hacia lo paradójico de la afirmación del escéptico —la insistencia en que lo que importa es las «cualidades sentidas incomunicables y especiales» de los estados internos. Pero si distinguimos entre la afirmación de que

> Tenemos acceso privilegiado a nuestros propios dolores

y la afirmación de que

> Sabemos en qué estados mentales estamos únicamente en virtud de sus cualidades sentidas especiales

[16] Cfr. P. F. Strawson, «Review of Wittgenstein's *Philosophical Investigations*» en la antología citada en la nota 13. En otra obra he señalado que podemos salvar las ideas epistemológicas de Wittgenstein, que se centran en torno a la imposibilidad de aprender el significado de las palabras sin una anterior «puesta en escena», sin verse atrapados en una hostilidad a lo privado que llevó a Wittgenstein al borde del conductismo y que hizo que algunos de sus seguidores franquearan ese límite. En la opinión que yo considero más acertada, la crítica de Wittgenstein de la «definición ostensiva pura» se puede generalizar a la doctrina de Sellars de que no podemos saber el significado de una palabra sin saber el significado de muchas otras y, por tanto, se puede utilizar para demostrar lo que hay de erróneo en la idea del Ojo de la Mente, sin extraer corolarios metafísicos. Véase mi «Wittgenstein, Privileged Access, and Incommunicability», *American Philosophical Quarterly,* 7 (1970), 192-205, y «Verificationism and Transcendental Arguments», *Nous,* 5 (1971), 3-14.

[17] Wittgenstein, *Philosophical Investigations,* pt. I, sec. 304.

podemos evitar la paradoja y admitir que una sensación es tan «algo» como una mesa. La primera afirmación dice únicamente que no hay mejor forma de averiguar si alguien tiene dolor que preguntárselo, y que nadie puede contrarrestar su propio informe sincero. La segunda dice que el mecanismo que hace posible este privilegio es su inspección de las «propiedades fenomenológicas» de sus propios estados mentales. Para pasar de la primera afirmación a la segunda necesitamos el modelo cartesiano del autoconocimiento en cuanto análogo a la observación —la imagen del Ojo Interior— y la idea de que los retortijones en el estómago, por ejemplo, no son Dados Naturalmente en la forma en que lo son las sensaciones producidas por dichos retortijones. Esta es la idea que, cuando se aplica a los Antípodas, produce la idea de que no pueden conocer directamente sus fibras-C sino que tienen que estar haciendo una inferencia «inconsciente» a partir de «cualidades sentidas, especiales».

Si abandonamos la idea de que la única forma en que podemos tener «conocimiento directo» de una entidad es conocer sus «cualidades incomunicables, sentidas, especiales», entonces podemos tener acceso privilegiado sin paradoja. Podemos abandonar esta idea si observamos, con Wittgenstein, que a no ser que haya eso que se llama conducta de dolor típica nunca podríamos enseñar a un niño el significado de «dolor de muelas», por ejemplo. Más en general, podemos advertir que la forma en que el niño prelingüístico sabe que tiene dolor es la forma en que el cambiadiscos sabe que el eje se ha quedado vacío, la planta sabe la dirección del sol, y la ameba la temperatura del agua. Pero esta forma no tiene conexión ninguna con lo que sabe el usuario del lenguaje cuando sabe qué es el dolor —que es mental más que físico, producido característicamente por lesión de algún tejido, etc. El error puesto al descubierto por Wittgenstein consistía en suponer que aprendemos lo que es un dolor en el segundo sentido poniendo un revestimiento lingüístico sobre nuestro conocimiento de lo que es el dolor en el primer sentido —revistiendo de palabras nuestro conocimiento directo de las cualidades incomunicables, sentidas y especiales (lo que nos volvería escépticos incorregibles en el tema de si se está denominando la misma cualidad incomunicable cuando nuestros amigos utilizan la misma palabra). La idea de que el conocimiento en el primer sentido —la clase manifestada por la discriminación conductual— es el «fundamento» (y no simplemente un posible antecedente causal) del conocimiento en el segundo sentido es en sí misma un producto más del modelo cartesiano. Mientras se piense que lo Naturalmente Dado se conoce gracias y sólo gracias a que es visto por el Ojo Interior, parecerá

extraño sugerir que la conducta y el entorno de que debemos tener conocimiento para *utilizar* la palabra «dolor» en la conversación normal deba tener algo que ver con lo que *significa* «dolor». La imagen del Ojo Mental, en combinación con la idea de que el lenguaje se compone de nombres de lo Naturalmente Dado más abreviaturas de criterios instrumentalistas para detectar la presencia de todas las realidades no Dadas Naturalmente, termina llevando al escepticismo. Estas suposiciones garantizan que los hechos sobre la conducta y el entorno que tan importantes son para los conductistas y wittgensteinianos parezcan irrelevantes a la «esencia» del dolor, pues tal esencia está determinada simplemente por lo que se *nombre*[18].

Desde el punto de vista de los Antípodas, la idea de «entidades privadas» —realidades sobre las que únicamente una persona tiene conocimientos incorregibles— es extraña pero no incomprensible. Para ellos es clara, pero inútil. Lo que les parecería ininteligible es la idea de entidades tan privadas, por así decirlo, que el conocimiento de las mismas no sólo sea privilegiado sino incomunicable. En realidad, éstas —podrían decir— no son entidades de ninguna clase. Por el contrario, la tentación wittgensteiniana de indicar que las sensaciones tienen alguna forma de existencia intermedia entre nadas y algos —que «salen» del mundo como los escarabajos metidos en cajas de la famosa analogía de Wittgenstein— viene de colocar juntas las ideas de incorregibilidad e incomunicabilidad. Si las asimilamos, y sacamos las implicaciones escépticas de la incomunicabilidad, seremos verdaderamente escépticos sobre el acceso privilegiado a las realidades privadas. Pero este escepticismo no será del mismo tipo que el que tienen los Antípodas. Estos son escépticos porque piensan que estas entidades y este acceso están *de trop*, no porque crean que las ideas de las mismas sean «confusiones conceptuales».

Los tradicionales escépticos cartesianos sobre las otras mentes son de una tercera categoría. Dudan de que las otras personas tengan,

---

[18] Podría discutirse si las opiniones de Wittgenstein sobre las sensaciones eran un corolario de sus puntos de vista sobre el significado, o si las ideas epistemológicas reflejadas en las primeras implican su filosofía del lenguaje. No creo que sirviera de mucho esta discusión. Como indicaré en el capítulo 6, no hay ninguna razón especial para considerar que la filosofía del lenguaje sea «anterior» a otras partes de la filosofía. En concreto, no creo que las disputas sobre la «teoría de la referencia» (de las que hablo en la sección 4 del capítulo 6) sirvan para iluminar la cuestión planteada por Kripke sobre la esencia del dolor, en el pasaje citado en la sección 2 de este capítulo. Las cuestiones sobre la designación rígida dejan sin aclarar las cuestiones sobre la esencia.

por ejemplo, dolores. Este escepticismo no es más refutable ni más interesante que el escepticismo sobre si existe la mesa cuando no se tiene delante ninguna que percibir. En último término, es perfectamente posible que las mesas desaparezcan cuano no hay nadie delante. Es muy posible que nuestros interlocutores simulen siempre la conducta de dolor, sin que nunca tengan dolores. Es igualmente posible que el mundo sea un lugar muy diferente a como lo imaginamos. Pero esta clase de escepticismo nunca habría merecido la atención de los filósofos si no fuera por la idea de lo Naturalmente Dado, y la consiguiente insinuación de que todo (la mente de nuestro amigo, así como su mesa y su cuerpo) lo que no es un fragmento de nuestro propio Espejo Interior —una parte de nuestra propia Esencia de Vidrio— sea simplemente un «postulado», «una inferencia», «una construcción», o algo igualmente dudoso que exija una construcción de sistemas metafísicos (Descartes, Kant) o descubrimientos sobre «nuestro lenguaje» (Russell, Ayer), si se quiere defender. Se supone que estas redescripciones de la realidad o del lenguaje demuestran que al escéptico le es *imposible* dudar de qué duda sin cometer un error intelectual general —«malinterpretar la lógica de nuestro lenguaje» (Ayer). Pero no es imposible, es simplemente útil, a no ser que se presente como razón de la duda algo que no sea la certeza de que no se puede tener.

No debemos hacernos la idea de que los filósofos del siglo XVII malinterpretaran la naturaleza de la mente distorsionando sistemáticamente, por así decirlo, el lenguaje corriente (Ryle). Tampoco hay que pensar que, dado que la metafísica ingenua del sentido común genera problemas de escepticismo, tengamos que sustituirla, por ejemplo, por un monismo neutro (Spinoza), o un materialismo (Smart), o un panpsiquismo (Whitehead, Hartshorne). El siglo XVII no «malinterpretó» el Espejo de la Naturaleza o el Ojo Interior más de lo que malinterpretara Aristóteles el movimiento natural o Newton la gravedad. Difícilmente *podían* malinterpretarlo, pues lo habían inventado ellos. La acusación de que este conjunto de imágenes inauguró una era de la filosofía que se centró en el escepticismo epistemológico es bastante válida, pero es importante darse cuenta de que esto *no* se debía a que otras mentes fueran especialmente sensibles al escepticismo. No son más sensibles que cualquier otra cosa que esté fuera de la propia mente. El escepticismo del siglo XVII dio nueva fuerza al escepticismo gracias a su epistemología, no a su filosofía de la mente. Toda teoría que entienda el conocimiento como exactitud de la representación, y que afirme

que sólo se puede estar razonablemente seguro sobre las representaciones, hará inevitable el escepticismo.

La epistemología del velo-de-ideas que se adueñó de la filosofía en el siglo XVII transformó el escepticismo de una curiosidad académica (escepticismo pirroniano) y una cuestión teológica concreta y local (la autoridad de la Iglesia frente a la del individuo que leía la Escritura) en una tradición cultural [19]. Lo hacía poniendo en pie un nuevo género filosófico —el sistema que vuelve a unir el sujeto y el objeto. Esta reconciliación ha sido el objeto del pensamiento filosófico desde entonces hasta ahora. Ryle y Wittgenstein están descaminados cuando dicen que tiene que haber algún error en la imagen del siglo XVII, que nos ha tenido cautivos, basándose en que en la vida corriente no tenemos ninguna dificultad para saber qué tiene una mente y qué no la tiene, ni si las mesas siguen existiendo cuando no están a la vista. Es como si dijéramos que la imitación de Cristo no puede ser un ideal aconsejable porque en la vida ordinaria no tenemos problemas para reconocer los límites impuestos al amor por la prudencia y el propio interés. Las imágenes que engendran tradiciones filosóficas (y poéticas) no es probable que se presenten fuera del estudio, lo mismo que los consejos de perfección que propone la religión no es probable que se aprecien en el trabajo diario. Si la filosofía es un intento de ver cómo «se integran lógicamente, en el sentido más amplio del término, las cosas, en el sentido más amplio del término», implicará siempre la construcción de imágenes que tendrán problemas característicos y engendrarán géneros escritos característicos. Alguien puede pensar, como es mi caso, que la imagen del siglo XVII es una imagen gastada —que la tradición que inspiró ha perdido vitalidad. Pero esta crítica es muy diferente a decir que esta tradición ha malinterpretado algo o no ha conseguido resolver un problema. El escepticismo y el género principal de la filosofía moderna tienen una relación simbiótica. No debe considerarse que la filosofía consiga su objetivo «dando respuesta al escéptico», ni que resulte ineficaz al comprobar que no hay ningún caso de escepticismo al que dar respuesta. La cosa es mucho más complicada.

---

[19] Véase Popkin, *Scepticism from Erasmus to Descartes,* y Maurice Mandelbaum, *Philosophy, Science, and Sense-Perception* (Baltimore, 1964), en los que se habla de varios factores que han contribuido a la formación de esta tradición. Véase una interpretación más radical en Jacques Maritain, *The Dream of Descartes* (Nueva York, 1944).

6. Materialismo sin identidad mente-cuerpo

Como el conductista y el escéptico sobre las otras mentes, el materialista tene una intuición válida que se vuelve paradójica cuando se formula en el vocabulario de la tradición a la que intenta reaccionar. Animado por el estudio de los Antípodas, el materialista cree probable que la referencia a los procesos y microestructuras neurológicas pueda reemplazar a la referencia a estados mentales de corta duración (sensaciones. pensamientos, imágenes mentales) en la explicación de la conducta humana. (Si es prudente, no opina lo mismo de las creencias, deseos y otros estados mentales de larga duración —y no cognoscibles incorregiblemente—, contentándose con considerarlos como propiedades de las personas más que de las mentes, a estilo de Ryle.) Sin embargo, no contento con esta predicción plausible, quiere decir algo metafísico. Parece que lo único que se puede decir es que «los estados mentales no son más que estados neurales». Pero esto parece pardaójico. Por eso intenta varias tácticas para reducir la paradoja. Una de estas tácticas es decir que la naturaleza de la mente ha sido malinterpretada hasta ahora, y que una vez que la comprendamos correctamente veremos que no resulta paradójico decir que puede resultar ser el sistema nervioso. El conductismo es una forma de esta táctica, y es compatible con el materialismo en el sentido de que la afirmación:

> Cuando hablamos de hechos mentales estamos hablando realmente de disposiciones conductuales

aunque no sea compatibles con:

> Cuando hablamos de hechos mentales estamos hablando realmente de hechos neurales

es compatible con:

> Hay, sin embargo, otras cosas relevantes para la predicción y explicación de la conducta que no son las interrelaciones sistemáticas de las disposiciones con los hechos del mundo exterior, y entre ellas están los hechos neurales que a veces causan el comienzo de tales disposiciones.

No obstante, en las discusiones de nuestros días lo habitual ha sido considerar conductismo y materialismo como dos formas totalmente diferentes —mitigada y violenta, respectivamente— de mo-

dificar la imagen de la mente del siglo XVII. En esta línea, los materialistas se han aferrado a las entidades mentales más recalcitrantes a los análisis disposicionales de Ryle —las sensaciones primarias, pensamientos pasajeros, imágenes mentales— y han tratado de demostrar que deben interpretarse, en términos generales, como «aquello, sea lo que sea, que *causa* el comienzo de cierta conducta o disposiciones conductuales». Este análisis de la mente, llamado análisis neutral con respecto al tema, y realizado especialmente por J. J. C. Smart y David Armstrong, tiene problemas con la distinción intuitiva entre «el estado mental, cualquiera que sea, que causa el comienzo...» y «el estado físico, cualquiera que sea, que causa el comienzo...». Dicho de otra manera, la intuición de que hay diferencia entre materialismo y paralelismo nos hace pensar que hay algo engañoso, o al menos incompleto, en las explicaciones neutras con relación al tema de lo que hace que lo mental sea mental. O, dicho con otras palabras, si nuestra noción de la «mente» es lo que dicen que es los análisis neutros en relación con el tema, es muy difícil explicar la existencia de un problema mente-cuerpo [20]. Podemos decir que la ausencia de una explicación neurológica de grano fino ha promovido la idea de que hay algo distintivo en la mente —que tiene que ser algo espritual—, pero esta táctica lo único que hace es dividir en dos partes la idea tradicional de lo mental: la función causal y la Esencia de Vidrio que se cree desempeña esta función causal. Evidentemente, los análisis neutros con relación al tema no pueden captar, ni quieren hacerlo, a esta última. Pero parece arbitrario dividir nuestro concepto de «estado mental» en la parte que es compatible con el materialismo y la parte que no, y luego decir que sólo la primera es «esencial» al concepto [21].

Podemos colocar en perspectiva el intento de realizar análisis «neutros con relación al tema» considerándolo como una forma de evitar el siguiente argumento en favor del dualismo:

1. Algunas afirmaciones de la forma «Acabo de tener una sensación de dolor» son verdaderas
2. Las sensaciones de dolor son hechos mentales
3. Los procesos neurales son hechos físicos

[20] Véase M. C. Bradley, «Critical Notice» de J. J. C. Smart, *Philosophy and Scientific Realism, Australasian Journal of Philosophy,* 42 (1964), 262-283. Mi «Incorrigibility as the Mark of the Mental», *Journal of Philosophy,* 67 (1970), 399-424, comienza con el estudio de este punto, y está inspirada en la revisión de Smart hecha por Bradley.

[21] Esto es lo contrario de la afirmación cartesiana de Kripke de que para aquél es esencial la «cualidad fenomenológica inmediata» de dolor.

4. «Mental» y «físico» son predicados incompatibles
5. Ninguna sensación de dolor es un hecho neural
6. Hay hechos no físicos

Los seguidores de Ryle, y algunos wittgensteinianos, que piensan que la realidad mental consiste en la accesibilidad al acceso privilegiado, y que dan rienda suelta a lo que Strawson llama «hostilidad a lo privado», rechazan el punto (2). Los panpsíquicos han rechazado (3)[22]. Los materialistas «reductivos» como Smart y Armstrong, que presentan análisis «neutros con relación al tema» de los términos mentales, se oponen a (4). Los materialistas «eliminativos» como Feyerabend y Quine niegan (1). Esta última postura proclama la ventaja, sobre la versión «reductiva», de no tener que ofrecer «análisis» revisionarios de los términos, y, por tanto, de no tener que complicarse con ideas dudosas como «significado» y «análisis». No dice que hayamos estado llamando erróneamente «sensaciones» a los procesos neurales, sino simplemente que no hay sensaciones. Tampoco dice que sea posible analizar el término *sensación* de forma que se produzcan resultados tan inesperados como la negación de (4). Es muy «quineano» y totalmente ant-ryleano en el sentido de que acepta alegremente todas las cosas que al dualista le gustaría interpretar que dice el hombre de la calle, y se limita a añadir: «Tanto peor para el hombre de la calle.»

Esta posición parece dar esperanzas a un sentido en que se puede comparar a bajo precio la afirmación metafísica de los materialistas: «Los estados mentales no son más que estados neurales.» Ahora es posible defenderla sin necesidad de hacer nada tan laborioso o sospechoso como el «análisis filosófico». Podemos decir que aunque en un sentido no haya sensaciones, en otro sentido lo que la gente *llamaba* «sensaciones», es decir, los estados neurales, existen realmente. La distinción no es más complicada que la que se da cuando decimos que no existe el cielo pero que hay algo a lo que la gente llama cielo (la apariencia de una cúpula azul como resultado

---

[22] Hartshorne y Whitehead son quizá los ejemplos más claros dentro de la filosofía de los últimos años. Yo he expuesto mis objeciones contra la versión de esta doctrina presentada por Whitehead en «The Subjectivist Principle and the Linguistic Turn» en *Alfred North Whitehead: Essays on His Philosophy,* ed. George Kline (Englewood Cliffs, N. J., 1963). Hay también una concepción panpsíquica en la propuesta de Thomas Nagel de una «fenomenología objetiva» que «haría posible que las cuestiones sobre la base física de la experiencia adoptaran una forma más inteligible» («What Is It Like to Be a Bat?», *Philosophical Review,* 83, 1974, 449). Sin embargo, tanto en Hartshorne como en Nagel, el panpsiquismo suele aparece ligado al monismo neutro.

de la refracción de la luz solar) que sí existe (aunque, como demuestra la difusión de la objeción de Brandt-Campbell expuesta en la sección 2, la analogía no puede llevarse hasta el punto de hacer de un estado mental la «apariencia» de un estado neural). Por tanto, parece que el argumento en favor del dualismo mencionado antes se puede resolver diciendo que a lo más que tiene derecho el dualista es a la siguiente premisa:

1'. Algunas afirmaciones de la forma «Acabo de tener una sensación de dolor» se consideran como verdaderas con la misma propiedad que «El cielo está nublado» y «Está saliendo el sol», pero ninguna de ellas *es* verdadera.

Si en el argumento utilizamos (1') en vez de (1), entonces sustituiremos (2) por lo siguiente:

2'. Si hubiera sensaciones de dolor, serían hechos mentales

y llegaremos a la conclusión:

6'. Las cosas que se han venido llamando «sensaciones» son hechos físicos (y, específicamente, neurales).

Podemos entonces concluir que aunque no haya hechos mentales, las cosas que se han llamado hechos mentales son hechos físicos, aun cuando «mental» y «físico» sean tan incompatibles como «que se levanta por encima del horizonte» y «que se queda quieto».

Este intento de lograr una versión barata de la identidad de mentes y cerebros funcionará bastante bien si nos abstenemos de hacer preguntas apremiantes sobre los criterios de identidad de referencia, de la misma manera que los análisis neutros con relación al tema funcionarán bien si nos abstenemos de hacer preguntas acuciantes sobre la identidad del significado. Sin embargo, no creo que haya criterios de la identidad de uno ni otro que sean útiles en los casos filosóficamente controvertidos. Por eso, no creo que el «materialismo eliminativo» sea una versión de la tesis de la identidad mente-cerebro que sea más plausible que el «materialismo reductivo». Cuando tratamos de dar sentido a cualquier afirmación con la forma «Realmente no hay Xs; de lo que venís hablando es únicamente de Ys», siempre es posible objetar que (a) «X» se refiere a Xs, y (b) no podemos referirnos a lo que no existe. Por ello, para salvar esta crítica clásica, el materialismo eliminativo tendría que decir que «sensación» no se refiere a la sensación, sino a nada en

absoluto, o que «referirse» en el sentido de «hablar sobre» no está sometido a (b). Cualquiera de las dos posturas es defendible, y yo defiendo la segunda al examinar la idea de referencia en conexión con el llamado problema del cambio conceptual (capítulo seis). Pero como creo que las versiones reductiva y eliminativa de la teoría de la identidad no son más que intentos torpes de introducir en la actual jerga filosófica nuestra reacción natural ante un encuentro con los Antípodas, no creo que se deba insistir en la diferencia entre las dos. Más bien, habría que abandonar ambas, y con ellas la noción de la «identidad mente-cuerpo». La reacción adecuada ante el caso de los Antípodas es adoptar un materialismo que no sea una teoría de la identidad en *ningún* sentido, y que evite de esta manera la idea artificial de que hemos de buscar «una teoría adecuada del significado (o referencia)» antes de decidir los problemas de la filosofía de la mente [23]. Esto equivale a decir, una vez más, que el materialismo debería dejar de reaccionar ante casos como el de los Antípodas diciendo cosas metafísicas, y limitarse a afirmaciones como «No se perdería ninguna capacidad de predicción, o explicación, o descripción, si hubiéramos hablado toda nuestra vida el lenguaje de los Antípodas». No tiene sentido preguntarse si el hecho de que los cerebroscopios corrijan los informes de los Antí-

[23] Esto no significa que las controversias en torno a las formas reductiva y eliminativa de la teoría de la identidad hayan resultado inútiles. Por el contrario, creo que su utilidad ha sido considerable, especialmente por su interacción con las cuestiones de la filosofía del lenguaje. Pero creo que el resultado de esta interacción ha sido, en primer lugar, reforzar la concepción de Quine de que no se puede invocar la idea de Quine de la «igualdad de significado» para resolver problemas filosóficos allí donde no ha servido la idea de «coextensivo», y, en segundo lugar, para demostrar que el sentido de «hablar realmente de» utilizado en discusiones como la del materialismo no tiene ninguna conexión interesante con la idea fregeana de referencia (en la que uno no se puede referir a lo que no existe). (El último punto se estudia en el capítulo sexto). Tras haber adoptado hace algunos años una posición materialista eliminativa («Mind-Body Identity, Privacy and Categories», *Review of Metaphysics,* 19 [1965], 25-54), estoy muy agradecido a las personas cuyas críticas a este artículo acabaron orientándome a lo que creo es una comprensión más clara de los problemas. Estoy especialmente agradecido a las publicaciones o conversaciones mantenidas con Richard Bernstein, Eric Bush, David Coder, James Cornman, David Hiley, William Lycan, George Pappas, David Rosenthal, Steven Savitt y Richard Sikora. El lector que tenga interés en profundizar en las semejanzas y diferencias entre materialismo reductivo y eliminativo podría consultar Cornman, *Materialism and Sensations* (New Haven, 1971); Lycan y Pappas, «What Is Eliminative Materialism?», *Australasian Journal of Philosophy,* 50 (1972), 149-159; Bush, «Rorty Revisited», *Philosophical Studies,* 25 (1974), 33-42, y Hiley, «Is 'Eliminative Materialism' Materialism?», *Philisophy and Phenomenological Research,* 38 (1978), 325-337.

podas sobre los estados internos demuestra que no son estados *mentales,* o demuestra más bien que los estados mentales son realmente estados neurales. Y no tiene sentido, no porque nadie tenga la menor idea de cómo resolver el problema, sino porque de él no depende nada. La sugerencia de que *tiene* una respuesta clara depende de la noción pre-quineana de las «condiciones necesarias y suficientes introducidas en nuestro lenguaje» para la aplicación de los términos «sensación», «mental», y otros parecidos, o de algún esencialismo semejante [24]. Sólo un filósofo muy interesado en la idea de «categoría ontológica» debería preocuparse de si un dolor del que se pueda informar incorregiblemente era «realmente» un dolor o más bien una fibra-C estimulada [25].

Si dejamos de hacer preguntas sobre lo que debe considerarse como «mental» y lo que no, y en vez de ello recordamos que en los enigmas relacionados con los Antípodas lo único que está en juego es la incorregibilidad, podremos ver que el argumento en favor del dualismo presentado más arriba es una versión más espectacular del siguiente argumento:

> 1. Algunas afirmaciones de la forma «He tenido una sensación de dolor» son verdaderas

[24] Yo cometí la imprudencia de acudir a esta idea en «Incorrigibility as the Mark of the Mental», citado en la nota 20, más arriba. Concluía allí que el desarrollo del debido respeto a los cerebroscopios significaría el descubrimiento de que nunca había habido hechos mentales. Pero esto es una exageración, y trata de establecer entre el materialismo eliminativo y reductivo una diferencia mayor de la que existe realmente (como han demostrado Lycan y Pappas). Me ha ayudado mucho a ver los fallos de mi anterior concepción la correspondencia mantenida con David Coder en relación con su «The Fundamental Error of Central State Materialism», *American Philosophical Quarterly,* 10 (1973), 289-298, y con David Rosenthal en torno a su «Mentality and Neutrality», *Journal of Philosophy,* 73 (1976), 386-415.

[25] Pero esto no equivale a decir que los Antípodas no producirían ningún impacto en la filosofía. La desaparición de la psicología como disciplina distinta de la neurología, y otros progresos culturales semejantes, podrían liberarnos con el tiempo de la imagen del Espejo de la Naturaleza mucho más eficazmente que las teorías de la identidad de los filósofos. Fuera de la filosofía, se produciría una cierta «confusión» en el habla normal (algunos casos de «no saber *qué* decir» cuando el introspector sincero desafiara al cerebroscopio), pero el sentido común, el lenguaje y la cultura han sobrevivido a otras confusiones peores que ésta. Compárense, por ejemplo, las conversaciones entre psiquíatras y jueces moralistas que presentan un historial clínico para demostrar que no es adecuado aplicar el término «criminal» a la conducta del acusado. Sólo un filósofo excesivamente celoso pensaría que se podría determinar una esencia del «crimen» examinando, por ejemplo, «nuestro lenguaje» y que sería capaz de resolver los dilemas de los jueces.

2'. Las sensaciones de dolor pueden ser objeto de informes incorregibles
3'. No se puede informar incorregiblemente de los hechos neurales
4'. No hay nada de lo que se pueda informar tanto corregible como incorregiblemente
5. Ninguna sensación de dolor es un hecho neural

En este caso, la tentación de evitar (5) negando (1) es mucho menor, pues es más fácil criticar (4') que (4). Es difícil decir que «mental» significa realmente «algo que podría resultar ser físico», lo mismo que cuesta decir que «conducta delictiva» significa realmente «conducta que podría resultar ser inocente». Esa es la razón por la que parecen condenados de antemano los intentos de realizar análisis neutros con relación al tema con la esperanza de negar (4'). Pero es relativamente fácil negar (4') y decir que es posible informar de algo corregiblemente (por los que saben neurología) e incorregiblemente (por los que no la saben), tan fácil como decir: «Algo puede ser tratado mejor que castigado (por los que entienden de psicología) y castigado más que tratado (por los que no)». En estos dos últimos ejemplos estamos hablando de prácticas sociales más que de «propiedades intrínsecas de las entidades en cuestión» o de «la lógica de nuestro lenguaje». Es fácil imaginarse diferentes prácticas sociales en relación con los mismos objetos, acciones o hechos, según cuál sea el grado de desarrollo intelectual y espiritual de la cultura en cuestión (siendo las fases «superiores» del desarrollo, con permiso de Hegel, aquellas en que el Espíritu es menos autoconsciente). Así, negando (4') parece que abrimos el camino a la negación de (5), y a decir que «sensación» y «procesos del cerebro» no son más que dos formas de hablar de la misma cosa.

Después de haber manifestado una actitud despectiva hacia los monistas neutrales y los teóricos de la identidad, quizá me esté metiendo ahora en su terreno. Efectivamente, se nos plantea la pregunta: ¿Dos formas de hablar de *qué?* ¿Algo mental o algo físico? Pero en este punto, creo yo, hemos de resistir nuestra tendencia metafísica natural, y *no* responder: «Una tercera cosa, de la que tanto la realidad mental como la realidad física son aspectos.» Sería mejor abandonar la argumentación y optar por el sarcasmo, haciendo preguntas retóricas como «¿De qué va este contraste mental-físico? ¿Quién dijo que todo lo que mencionaba alguien debía pertenecer a uno u otro de los dos (o media docena de) dominios ontológicos?» Pero esta táctica no parece muy ingeniosa, pues resulta obvio (una vez que los departamentos de psicología dejen de hacer experimentos

con cuestionarios y diapositivas, y trabajen únicamente con cerebroscopios) que de alguna manera ha triunfado «lo físico».

Pero, ¿sobre qué triunfó? ¿Sobre lo mental? ¿Qué era eso? ¿La práctica de hacer informes incorregibles sobre algunos de los propios estados? Eso parece algo demasiado diminuto para figurar como una revolución intelectual. Entonces, quizá triunfara sobre la convicción del intelectual sentimental de que había un dominio interno privado en el que no podían penetrar la publicidad, «los métodos científicos» ni la sociedad. Pero tampoco esto es exacto. El secreto del corazón del poeta permanece oculto para la policía secreta, a pesar de su capacidad de predecir cada uno de sus pensamientos, elocuciones y movimientos examinando el cerebroscopio que él ha de llevar día y noche. Podemos saber qué pensamientos cruzan la mente de un hombre sin entenderlos. Nuestra singularidad inviolable estriba en nuestra capacidad de decirnos cosas obvias nada más que a nosotros.

La verdadera dificultad con que nos encontramos aquí es, una vez más, que estamos tratando de dejar de lado la imagen del hombre en cuanto poseedor de una Esencia de Vidrio adecuada para reflejar la naturaleza, por una parte, y por la otra estamos aferrándonos a ella. Si pudiéramos llegar a abandonar todo el cúmulo de imágenes que los Antípodas no comparten con nosotros, no seríamos capaces de inferir que la materia había triunfado sobre el espíritu, la ciencia sobre lo privado, ni ninguna cosa sobre otra. Estos contrarios enfrentados son ideas que no tienen sentido fuera de un conjunto de imágenes heredadas del siglo XVII de nuestro planeta. Nadie, excepto los filósofos, que están profesionalmente obligados a tomarse en serio en estas imágenes, se escandalizará si la gente comienza a decir: «La máquina me dijo que no hacía realmente daño —lo único que parecía que lo hacía, y además terriblemente.» Los filósofos están demasiado ocupados en ideas como «categoría ontológica» para tomarse a la ligera estos acontecimientos, pero ningún otro sector de la cultura lo está. Pensemos en el hecho de que *sólo* los filósofos siguen perplejos ante la cuestión de cómo se pueden tener motivos y deseos inconscientes). Unicamente la idea de que la filosofía debe proporcionar una matriz permanente de categorías en que se pueda encajar sin violencia todo posible descubrimiento empírico y desarrollo cultural nos mueve a hacernos preguntas sin respuesta como «¿Significaría esto que no había mentes?». «¿Estábamos equivocados sobre la naturaleza de la mente?» «¿Tenían razón los Antípodas cuando decían: 'Nunca

ha habido ninguna de esas cosas que vosotros llamabais *sensaciones primarias?*»

Finalmente, la misma concepción, excesivamente ambiciosa, de la filosofía, procedente del mismo conjunto de imágenes del siglo XVII, es responsable de los temores de los materialistas a que, si la localización cerebral y el «análisis filosófico» no cooperan en la tarea de «identificar» la mente y el cuerpo, se pone en peligro «la unidad de la ciencia». Si seguimos a Sellars cuando dice que la ciencia es la medida de todas las cosas, entonces no tendremos que preocuparnos de que la localización cerebral resulte ser un fracaso, y mucho menos de que los «análisis» materialistas de nuestro vocabulario mentalista cotidiano sucumban a los ejemplos contrarios. No interpretaremos ninguna de las dos cosas como demostración de que la ciencia haya estado montada todo ese tiempo sobre dos caballos —uno sólido y otro fantasmagórico— que pueden comenzar a galopar en direcciones contrarias en cualquier momento. El que la ciencia no llegue a imaginar cómo funciona el cerebro no será mayor peligro para la «unidad» de la ciencia que el que no llegara a explicar la mononucleosis, o la migración de las mariposas o los ciclos del mercado bursátil. Aunque resultara que las neuronas «se escapaban» —zarandeadas por fuerzas aún desconocidas para la ciencia— no se reivindicaría a Descartes. Pensar lo contrario sería incurrir en la falacia de *omne ignotum pro spectro* —suponer que todo lo que no se puede entender es un fantasma, algo que desde antes sabemos que está fuera del alcance de la ciencia, y, por lo tanto, se debe poner, desesperadamente, en manos de la filosofía [26].

[26] Se puede precisar un poco más este punto utilizando la distinción de Meehl y Sellars entre «físico$_1$» («un hecho o entidad es físico$_1$ si está dentro de la red del espacio-tiempo») y «físico$_2$» («un hecho o entidad es físico$_2$ si se puede definir en términos de primitivos teóricos adecuados para describir completamente los estados reales, aunque no necesariamente las posibilidades del universo antes de la aparición de la vida»). Esta distinción (expuesta en «The Concept of Emergence», *Minnesota Studies in the Philosophy of Science,* I [1956], 252) se puede multiplicar distinguiendo «físico$_1$» de los sentidos adecuados de «físico$_n$» definidos en términos de «el universo antes de la aparición de la conducta lingüística» («... de la acción intencional», «de las creencias y deseos», etc.). En cualquiera de estas distinciones, lo que importa subrayar es que el que la ciencia no consiga explicar algo en términos de entidades físicas$_n$ (siendo *n* mayor que *1*) no demuestra en absoluto que la explicación deba hacerse en términos de entidades físicas$_1$. Geoffrey Hellman y Frank Thompson sacan buen partido de este punto en «Physicalism: Ontology, Determination and Reduction», *Journal of Philosophy,* 72 (1975), 551-564, y «Physicalist Materialism», *Nous,* 11 (1977), 309-346. Su versión del «materialismo sin identidad» debe compararse con la de Davidson en su «Mental Events» (de la que nos ocupamos en el capítulo 4, más adelante).

Si no creemos que la filosofía proporcione un marco ontológico permanente para ningún posible resultado científico (que consistiera, por ejemplo, en categorías como «mental» y «físico») no pensaremos que el fallo de la ciencia sea una reivindicación de Descartes, lo mismo que no consideramos como reivindicación de Tomás de Aquino el que la ciencia no consiga explicar el origen de la primera célula viva. Si las neuronas se nos escapan, o si el cerebro funciona holísticamente más que atomísticamente, esto no nos ayuda a demostrar que tengamos, en definitiva, ideas claras y distintas de «lo mental» y «lo físico». Estas categorías llamadas ontológicas no son más que formas de agrupar ideas más bien heterogéneas, procedentes de fuentes históricas bastante diversas, que convenían a los intereses de Descartes. Pero sus intereses no son los nuestros. Los filósofos no deberían pensar que este conglomerado artificial sea un descubrimiento de algo preexistente —descubrimiento que, por ser «intuitivo» o «conceptual» o «categórico», establece parámetros permanentes para la ciencia y la filosofía.

7. Epistemología y «la filosofía de la mente»

Espero que los dos capítulos sobre el «problema mente-cuerpo» que el lector acaba de recorrer le hayan convencido al menos de los siguientes puntos:

> Si no queremos revivir las ideas platónicas y aristotélicas sobre la comprensión de los universales, no pensaremos que el conocimiento de las verdades generales sea posible gracias a un componente especial, metafísicamente distintivo, de los seres humanos.
>
> Si no queremos resucitar el uso, torpe y algo incoherente, que el siglo XVII hizo de la idea aristotélica de «sustancia» no veremos ningún sentido a la idea de los dos reinos ontológicos —el mental y el físico.
>
> Si no queremos afirmar lo que he llamado Principio (P) —en términos generales, la afirmación de que una propiedad metafísica distintiva de la «presencia a la conciencia»» sirve de base a algunos de los informes no inferenciales de nuestros estados— no podremos utilizar la idea de «entidades cuya apariencia agota su realidad» para reforzar la distinción mental-físico.

La idea de que hay un problema sobre la mente y el cuerpo se originó en el intento del siglo XVII de convertir a «la mente» en

una esfera independiente de investigación. La idea era ofrecer una explicación para-mecánica de los procesos mentales que, de alguna manera, confirmarían algunas pretensiones de conocimiento y descalificarían a otras. El paradigma del «giro epistemológico» adoptado por la filosofía en el siglo XVII era lo que Kant llamó «la fisiología de la comprensión humana del famoso Mr. Locke» —una explicación causal de los procesos mentales que se supone sirve para criticar y justificar las pretensiones de conocimiento. Para despegar del suelo, esta idea necesitó la sustitución cartesiana del problema antiguo y medieval de la razón por el problema moderno de la conciencia. Si lo que he dicho en los dos capítulos anteriores es cierto, la persistencia de ideas como el «problema mente-cuerpo» y «la filosofía de la mente» se debe a la persistencia de la idea de que hay alguna conexión entre las antiguas nociones de razón o personalidad y la noción cartesiana de conciencia. En la parte II del libro se intenta destruir la versión moderna del problema de la razón —la idea de que hay un problema en relación con la posibilidad o amplitud de la representación precisa que es el tema de que se ocupa una disciplina llamada «epistemología». En la medida en que logre su objetivo, este intento nos libra de la idea del conocimiento humano en cuanto reunión de representaciones en un Espejo de la Naturaleza, y, por tanto, confirma la afirmación de la parte I de que podemos prescindir de la idea de nuestra Esencia de Vidrio. Si el conocimiento no consiste en la exactitud de las representaciones, como no sea en el sentido más trivial y no problemático, tampoco necesitamos ningún espejo interior y, por lo tanto, no hay ningún misterio sobre la relación de ese espejo con nuestros miembros más bastos.

Sin embargo, aunque destruyamos los problemas de la conciencia y la razón, podría seguir intacto el de la personalidad, pues esta noción se basa en nuestras intuiciones morales, intuiciones que no es probable que sean simplemente resultado de los desviados intentos de los griegos o del siglo XVIII por construir modelos del conocer o de la mente. El tema de la personalidad lo consideraremos en la tercera parte del libro, en conexión con la idea de «filosofía». Allí, trato de hacer ver cómo el proyecto peculiarmente filosófico de descubrir qué entidades son personas, y, por lo tanto, poseen dignidad moral, basándose en algún «criterio objetivo» —por ejemplo, su posesión de una Esencia de Vidrio— es una confusión, aproximadamente, entre ciencia y ética. En la parte III trato de sugerir una forma de considerar nuestra conciencia moral en que se evita esta idea.

Así pues, en las partes II y III se consderan las ideas que he enumerado en la sección 3 del capítulo como «señales de lo mental» y no han sido examinadas en la parte I. Se sugieren formas de utilizar estas ideas —por ejemplo, la intencionalidad, la dignidad moral— una vez que se ha abandonado la idea de un ingrediente mental especial —nuestra Esencia de Vidrio. Por ello, tengo la esperanza de que en las páginas que quedan se disipen algunas de las dudas que se haya planteado el lector a lo largo de la parte I y algunas de las cuestiones que hayan podido quedar en el aire.

SEGUNDA PARTE

El reflejo

Capítulo III



La idea de una "Teoría del Conocimiento"

1. Epistemología y auto-imagen de la filosofía

La idea de que existe una disciplina autónoma llamada «filosofía», distinta de la religión y de la ciencia y capaz de emitir juicios sobre ambas, es de origen muy reciente. Cuando Descartes y Hobbes acataban «la filosofía de las escuelas» no pensaban que ellos estaban sustituyéndola por una filosofía nueva y mejor —una mejor teoría del conocimiento, o una metafísica mejor, o una ética mejor. Estas distinciones entre «campos de la filosofía» no habían llegado a trazarse todavía. No se disponía todavía de la idea de la «filosofía» misma, en el sentido en que se ha venido entendiendo desde que el tema se consagró en forma de materia académica durante el siglo XIX. Mirando hacia atrás, vemos a Descartes y a Hobbes como «iniciadores de la filosofía moderna», pero ellos pensaban en su función cultural en términos de lo que Lecky llamaría «la guerra entre la ciencia y la teología». Estaban luchando (aunque discretamente) para conseguir que el mundo intelectual fuera seguro para Copérnico y Galileo. No se veían a sí mismos como si estuvieran ofreciendo «sistemas filosóficos», sino como contribuidores al florecimiento de la investigación en matemáticas y mecánica, y como liberadores de la vida intelectual frente a las instituciones eclesiásticas. Hobbes definía la «filosofía» como «aquel conocimiento de los efectos de las apariencias que adquirimos por raciocinio verdadero a partir del conocimiento que tenemos previamente de las causas de su generación» [1]. No tenía ningún interés en distinguir lo que estaba haciendo de una cosa diferente llamada «ciencia». Sólo des-

---

[1] Thomas Hobbes, *De Corpore*, cap. I, sec. 2.

pués de Kant se impuso la moderna distinción filosofía-ciencia. Hasta que no se quebró del dominio de las iglesias sobre la ciencia y la erudición, las energías de los hombres a quienes ahora consideramos como «filósofos» se dirigían a la demarcación de sus actividades separándolas de la religión. Sólo cuando se hubo ganado esa batalla pudo plantearse la cuestión de la separación de las ciencias.

La demarcación entre filosofía y ciencia que llegó a imponerse fue posible gracias a la idea de que el núcleo de la filosofía era la «teoría del conocimiento», una teoría distinta de las ciencias debido a que era su *fundamento*. Ahora podemos hacer remontar esa idea al menos hasta las *Meditaciones* de Descartes y a *De Emendatione Intellectus,* pero no consiguió la auto-conciencia hasta Kant. No se incorporó a la estructura de las instituciones académicas, y a las auto-descripciones espontáneas de los profesores de filosofía, hasta bien entrado el siglo XIX. Sin esta idea de una «teoría del conocimiento», es difícil imaginar qué «filosofía» podría haber habido en la ciencia moderna. La metafísica —considerada como la descripción de cómo se encajan los cielos y la tierra— había sido desplazada por la física. La secularización del pensamiento moral, que fue la preocupación dominante de los intelectuales europeos de los siglos XVII y XVIII, no se consideraba entonces como la búsqueda de un nuevo fundamento metafísico que ocupara el lugar de la metafísica teísta. Kant, sin embargo, consiguió transformar la antigua idea de la filosofía —la metafísica en cuanto «reina de las ciencias» por ocuparse de lo que era más universal y menos material— en la idea de la disciplina «más básica» —una disciplina con carácter de *fundamento*. La filosofía se convirtió en «primaria» no ya en el sentido de «la más alta» sino en el sentido de «subyacente». Cuando Kant hubo escrito su obra, los historiadores de la filosofía pudieron situar a los pensadores de los siglos XVII y XVIII como hombres que trataban de dar respuesta a la pregunta «¿Cómo es posible nuestro conocimiento?» e incluso de proyectar esta cuestión hasta los pensadores de la antigüedad[2].

---

[2] Sobre la diferencia entre historias de la filosofía escritas antes y después de Kant, véase Maurice Mandelbaum, «On the Historiography of Philosophy», *Philosophy Research Archives,* vol. II (1976); John Passmore, «Historiography of Philosophy», *Philosophy Research Archives,* vol. II (1976); John Passmore, «Historiography of Philosophy» en *The Encyclopedia of Philosophy* (Nueva York, 1967); Lucien Braun, *Histoire de l'Histoire de la Philosophie* (París, 1973), especialmente el capítulo 5; Victor Cousin, *Introduction à l'Histoire de la Philosophie* (París, 1868), douzième leçon, «Des Historiens de la Philosophie». Mandelbaum (pág. 713) dice, hablando de la frecuencia con que se

No obstante, esta imagen kantiana de la filosofía en cuanto centrada en la epistemología sólo consiguió aceptación general cuando Hegel y el idealismo especulativo dejaron de dominar el panorama intelectual de Alemania. La filosofía sólo llegó a profesionalizarse del todo después de que hombres como Zeller comenzaran a decir que ya era hora de dejar de edificar sistemas y dedicarse a la labor paciente de separar lo «dado» de las «adiciones subjetivas» realizadas por la mente [8]. El movimiento de «vuelta a Kant» producido después de 1860 en Alemania fue también un movimiento de «vamos a ponernos a trabajar» —una forma de separar la disciplina autónoma y no empírica que era la filosofía de la ideología, por una parte, y de la naciente ciencia de la psicología empírica, por la otra. La imagen de la «epistemología-y-metafísica» como «centro de la filosofía» (y de la «metafísica»» como algo que surge de la epistemología, y no al revés), que establecieron los neo-kantianos,

---

estudió el concepto de historia de la filosofía durante la última década del siglo XVIII en Alemania, que «creo que lo que ocurre —aunque mi documentación sea insuficiente para que esto pase de ser una mera conjetura— es que esta discusión tuvo que estar instigada por el impacto de la obra de Kant en aquel periodo: por la sensación de que su sistema era al mismo tiempo un final y nuevo comienzo». Todos estos escritores insisten en el contraste entre Brucker, *Historia Critica Philosophiae,* 1742-1767, y las historias de Tiedeman (*Geist der spekulativen Philosophie,* 1791-1797) y Tennemann (*Geschichte der Philosophie,* 1789-1819). Brucker dedica aproximadamente la décima parte de sus páginas a los «filósofos sincréticos» modernos (es decir, los que no encajan en una de las antiguas escuelas) e incluye (además de la secuencia ahora ya clásica formada por Bacon, Descartes, Hobbes, Spinoza, etc.) más de veinte figuras distintas (por ejemplo, Maquiavelo, Kepler, Boyle). Éstos no son tenidos en cuenta por Tiedemann, que origina la breve lista canónica de «grandes filósofos modernos». Tennemann nos dice que «la historia de la filosofía es la presentación de sucesivas etapas en la evolución de la filosofía, la presentación de los esfuerzos de la Razón por realizar la idea de una ciencia de las leyes y bases últimas de la Naturaleza y de la Libertad», después de habernos dicho previamente que la razón comienza con la facultad de «unificar la diversidad de las representaciones», y de ahí pasa a la unificación última del pensamiento científico (*Geschichte der Philosophie,* vol. I, Leipzig, 1798, págs. XXIX, XXVI). Su historia de la filosofía tiene la cualidad «dramática» que asociamos con Hegel. Nos presenta un canon de los que deben figurar como filósofos, basado en el mayor o menor parecido de su obra con la de Kant, y la noción de progreso en la filosofía desde la antigüedad hasta los autores modernos.

[8] Cfr. Edouard Zeller, «Uber Bedeutung und Aufgabe der Erkenntnistheorie», en *Vorträge und Abhandlungen,* Zweite Sammlung (Leipzig, 1877), página 495. El ensayo de Zeller forma parte de una larga secuencia procedente de Kant que anuncia que se ha acabado la época del filósofo amateur y que ahora les toca a los profesionales. Véase una versión posterior en G. J. Warnock, *English Philosophy Since 1900* (Londres, 1958), pág. 171: «... la filosofía ha conseguido categoría profesional en los últimos años» y pág. 172: «... has-

es la que reflejan los programas de filosofía de la actualidad [4]. La misma expresión *teoría del conocimiento* consiguió credibilidad y respetabilidad sólo cuando Hegel quedó caduco. La primera generación de admiradores de Kant utilizaba el término *Vernunftkritik* como una forma cómoda de referirse a «lo que hizo Kant», y las palabras *Erkenntnislehre* y *Erkenntnistheorie* se inventaron algo más tarde (en 1808 y 1832, respectivamente) [5]. Pero para entonces Hegel y la creación de sistemas idealistas habían logrado oscurecer la pregunta «¿Cuál es la relación de la filosofía con las otras disciplinas?». El hegelianismo produjo una imagen de la filosofía como disciplina que de alguna manera completaba y se tragaba a las demás disciplinas, en vez de servirles de *base*. Además, hizo a la filosofía demasiado popular, demasiado interesante, demasiado importante, para ser propiamente profesional; estimuló a los profesores de filosofía a encarnar el Espíritu del Mundo, en vez de limitarse a continuar con su *Fach*. El ensayo de Zeller que (según Mauthner) «fue el primero que elevó el término 'Erkenntnistheorie' a su actual

ta hace muy poco no se ha distinguido claramente el objeto material, o más bien las tareas, de la filosofía de las de otras disciplinas» (Warnock está hablando del periodo 1900-1968). No sé cuándo aparecerían estas afirmaciones por primera vez, pero véase el contraste del uso del término *Philosophen von Profession,* de Ernst Reinhold, con *alle wissenschaftlich Gebildete* a quien estaba destinado su *Handbuch der allgemeinen Geschichte der Philosophie*. (*Handbuch...*, Erster Theil, Gotha, 1828, pág. v). El *Einleitung* a la historia de Reinhold es útil para confirmar la afirmación de Mandelbaum (cfr. nota 2, más arriba) sobre la relación del pensamiento kantiano con la sustitución de «las crónicas de las opiniones de los filósofos» por «la historia de la filosofía» en el sentido moderno.

[4] Como trataré de demostrar más detalladamente en páginas posteriores, especialmente en el capítulo 6, la idea contemporánea de filosofía del lenguaje en cuanto «filosofía primera» no significa un cambio tan grande de la afirmación anterior de que la epistemología era «primera», sino más bien una pequeña variante de la misma. La afirmación central de la filosofía desde Kant hasta nuestros días ha sido que la que necesitaba explicación era la «posibilidad de representar la realidad», y para este proyecto la diferencia entre representaciones mentales y lingüísticas es relativamente irrelevante.

[5] Cfr. Hans Vaihinger, «Über den Ursprung des Wortes 'Erkenntnistheorie'», *Philosophische Monatshefte,* vol. XII (Leipzig, 1876), págs. 84-90, sobre la historia del término. Vaihinger presenta el punto de vista, que debían compartir la mayoría de los neo-kantianos, y que yo estoy adoptando en este lugar, de que Locke fue el primero en «tener una conciencia clara de que toda discusión metafísica y ética ha de ir precedida de investigaciones epistemológicas» y que lo que Descartes y Spinoza dijeron en este sentido era meramente ocasional y asistemático (pág. 84). Debo la referencia a Vaihinger (así como otras referencias y muchas ideas iluminadoras) a un artículo sin publicar de Ian Hacking en que estudia la aparición de la epistemología como disciplina.

dignidad académica»[6], termina diciendo que los que creen que podemos extraer todas las ciencias de nuestro propio espíritu pueden seguir con Hegel, pero que los más sensatos deberían reconocer que la tarea propia de la filosofía (una vez rechazada la idea de la cosa-en-sí y, por tanto, las tentaciones de idealismo) es establecer la objetividad de las pretensiones de conocimiento de las distintas disciplinas empíricas. Esto se conseguirá mediante la adecuación de las aportaciones a priori aplicadas en la percepción[7]. Así pues, la *Erkenntnistheorie* aparece en 1862 como una forma de eludir el «idealismo» y la «especulación». Quince años más tarde, Zeller observa que ya no es necesario señalar el papel adecuado de la *Erkenntnistheorie,* pues ya ha sido aceptado en general, especialmente por parte de «nuestros colegas más jóvenes»[8]. Otros treinta años después, William James se quejaría «del temperamento plúmbeo de nuestros jóvenes y calvos doctores en filosofía, que se dedican a aburrirse mutuamente en los seminarios, que escriben esos terribles informes bibliográficos en la *Philosophical Review* y en otras partes, y se alimentan de 'obras de referencia' y nunca confunden 'Aesthetik' con 'Erkenntnistheorie'»[9].

En el presente capítulo quiero rastrear algunas de las etapas cruciales en la transición de las campañas de Descartes y Hobbes contra «la filosofía de las escuelas» al restablecimiento decimonónico de la filosofía en cuanto disciplina autónoma, independiente, «escolástica». Trataré de confirmar la afirmación (común a Wittgenstein y Dewey) de que pensar en el conocimiento que presenta un «problema», y sobre el que deberíamos tener una «teoría», es resultado de considerar el conocimiento como una agrupación de representaciones —concepción del conocimiento que, como vengo señalando, fue producto del siglo XVII. La consecuencia que se debe extraer es que si esta forma de pensar en el conocimiento es optativa, también lo es la epistemología, y también lo es la filosofía tal como se ha entendido a sí misma desde mediados del siglo pasado. La

[6] Fritz Mauthner, *Wörterbuch der Philosophie* (Munich y Leipzig, 1910), sección V, «Erkenntnistheorie», vol. I, pág. 296: «ein ausschliesslich deutscher Ausdruck, von Reinhold der Sohne (nach Eisler) geprägt, aber erst durch Zeller zu seiner jetzigen akademischen Würde promoviert». Vaihinger dice lo mismo de Zeller: «Wortes 'Erkenntnistheorie'», pág. 89. El artículo de Vaihinger es un informe sobre (y un ejemplo de) la nueva auto-imagen «profesionalizada» que los filósofos neo-kantianos estaban creando.

[7] Zeller, «Erkenntnistheorie», págs. 494-495.

[8] *Ibíd.*, pág. 496.

[9] William James, *Letters,* ed. Henry James (Boston, 1920), pág. 228 (carta a George Santayana, del 2 de mayo de 1905).

historia que voy a relatar sobre la forma en que la filosofía-como-epistemología adquirió certeza de sí misma en el periodo moderno seguirá estas líneas:

La invención de la mente hecha por Descartes —su fusión de creencias y sensaciones en las ideas lockeanas— dio a la filosofía nueva base en que apoyarse. Ofreció un campo de investigación que parecía «previo» a los temas sobre que se habían manifestado los antiguos filósofos. Además, proporcionaba un ámbito dentro del cual era posible la *certeza,* en oposición a la mera *opinión.*

Locke convirtió la «mente» de nuevo diseño de Descartes en el objeto material de una «ciencia del hombre» —la filosofía moral en cuanto opuesta a la filosofía natural. Lo hizo pensando confusamente que una analogía de la mecánica de partículas de Newton aplicada al «espacio interior» supondría de alguna manera «una gran ventaja en la dirección de nuestros Pensamientos en la búsqueda de otras Cosas» [10] y nos dejaría «ver de qué objetos se podía ocupar adecuadamente nuestra Comprensión, y de cuáles no» [11].

Este proyecto de aprender más sobre lo que podíamos saber y cómo podíamos saberlo mejor estudiando cómo funciona nuestra mente llegaría con el tiempo a ser bautizado con el nombre de «epistemología». Pero antes de que el proyecto llegara a la auto-conciencia total, había que encontrar una forma de convertirlo en un proyecto *no-empírico.* Tenía que ser algo conseguido en la reflexión reposada, independiente de los descubrimientos fisiológicos y capaz de producir verdades necesarias. Locke, que había conservado el nuevo espacio interior de la investigación —producto de la mente recién inventada por Descartes—, no había podido mantenerse fiel a la certeza cartesiana. El «sensualismo» de Locke no era todavía un candidato adecuado para ocupar la posición vacante de la «reina de las ciencias».

Kant colocó a la filosofía «en el camino seguro de la ciencia» colocando el espacio exterior dentro del espacio interior (el espacio de la actividad constituyente del ego transcendental) y afirmando luego la certeza cartesiana sobre el interior para las leyes de lo que antes se había considerado como exterior. De esta manera reconcilió la afirmación cartesiana de que podemos tener certeza únicamente de nuestras ideas, con el hecho de que ya teníamos certeza —conocimiento a priori— de lo que parecían no ser ideas. La revolución copernicana se basaba en la idea de que sólo podemos tener conocimiento de los objetos a priori si los «constituimos», y Kant no

[10] John Locke, *Essay,* I, I, i.
[11] *Ibid.,* «Epistle to the Reader».

se preocupó nunca por la pregunta de cómo podríamos tener conocimiento apodíctico de estas «actividades constituyentes», pues se suponía que eso corría a cargo del acceso privilegiado cartesiano [12]. Una vez que Kant hubo sustituido la «filosofía de la comprensión humana del celebrado Mr. Locke» por (en palabras de Strawson) «la materia mítica de la psicología transcendental», la «epistemología» había llegado a la mayoría de edad.

Además de elevar «la ciencia del hombre» de un nivel empírico a un nivel a priori, Kant hizo otras tres cosas que contribuyeron a que la filosofía-como-epistemología adquiriera confianza en sí misma y carácter auto-consciente. En primer lugar, al identificar como problema central de la epistemología las relaciones entre dos clases de representaciones —«formales» (conceptos) y «materiales» (intuiciones)— igualmente reales pero irreductiblemente distintas, hizo posible que se vieran importantes continuidades entre la nueva problemática epistemológica y los problemas (los problemas de la razón y de los universales) que habían preocupado a los hombres de la Antigüedad y de la Edad Media. De esta manera hizo posible que se escribieran «historias de la filosofía» de corte moderno. En segundo lugar, al unir la epistemología con la moralidad en el proyecto de «destruir la razón para dejar lugar a la fe» (es decir, destruir el determinismo newtoniano para dejar sitio a la conciencia moral común), revivió la idea de un «sistema filosófico completo», en el que la moralidad tenía como «base» algo más científico y menos polémico. Mientras que las antiguas escuelas tenían distintas concepciones de la virtud humana, destinadas a adaptarse a su concepción de cómo era el mundo, Newton se había adueñado de las concepciones en este último tema. Con Kant, la epistemología pudo ocupar el papel de la metafísica en cuanto garantía de los presupuestos de la moralidad. En tercer lugar, al suponer que todo lo que decimos trata sobre algo que hemos «constituido», hizo posible que la epistemología fuera considerada como ciencia de carácter básico, una disciplina basada en la reflexión y capaz de descubrir las características «formales» (o, en versiones posteriores, «estructurales», «fenomenológicas», «gramaticales», «lógicas», o «con-

---

[12] Véase *K.d.r.V.*, Bxvi-xvii: La suposición de que «los objetos deben conformarse a nuestro conocimiento», dice Kant, «encaja mejor con la suposición de que sería posible tener conocimiento de objetos *a priori* determinando algo en relación con ellos antes de que sean dados». La cuestión de cómo sabemos a qué condiciones deben conformarse —cómo dar validez a las pretensiones de conocimiento hechas desde el punto de vista trascendental— no se trata aquí ni en ninguna otra parte de la primera *Crítica*.

ceptuales») de cualquier área de la vida humana. De esta manera hizo posible que los profesores de filosofía se vieran a sí mismos como si estuvieran presidiendo un tribunal de la razón pura, capaces de determinar si otras disciplinas estaban dentro de los límites legales fijados por la «estructura» de sus objetos materiales»[13].

2. La confusión de Locke entre explicación y justificación

El «giro epistemológico» realizado por Descartes quizá no se habría adueñado de la imaginación de Europa si no hubiera sido por una crisis de confianza en las instituciones establecidas, crisis expresada paradigmáticamente en Montaigne. Pero deberíamos distinguir el escepticismo pirroniano tradicional sobre nuestra capacidad de llegar a la certeza del nuevo escepticismo del velo-de-ideas que hizo posible Descartes al labrar el espacio interior. El escepticismo tradicional se había inquietado principalmente por el «problema del criterio» —el problema de dar validez a los procedimientos de investigación y evitar al mismo tiempo la petición de principios o el dogmatismo. Este problema, que Descartes pensaba que había resuelto él mismo con «el método de las ideas claras y distintas», tenía poco que ver con el problema de pasar del espacio interior al espacio exterior —el «problema del mundo exterior» que adquirió carácter paradigmático para la filosofía moderna[14]. La idea de

[13] Ha sido muy comentada la afición de Kant a las metáforas de la jurisprudencia. También se observa en los neo-kantianos. Véase, por ejemplo, Zeller, «Über die Aufgabe der Philosophie und ihre Stellung zu den übrigen Wissenschaften» (*Vorträge und Abhandlungen,* Zweite, Sammlung, págs. 445-466). Al trazar las pretensiones del filósofo como supervisor de la cultura, habla de la *Rechtstitel* que cada disciplina debía conseguir de la filosofía y dice que «no hay ninguna rama del conocimiento humano cuyas raíces no profundicen hasta el terreno de la filosofía, pues toda ciencia procede del espíritu cognoscente y toma sus leyes de los procedimientos de ese espíritu» (pág. 465).

[14] Las difíciles relaciones entre estas dos formas de escepticismo aparecen ilustradas por la sección XII de Hume, *Enquiry Concerning Human Understanding,* «Of the Academical or Sceptical Philosophy». Hume quiere distinguir el escepticismo de la *Primera Meditación* de Descartes (que le parecía hiperbólico e imposible) del escepticismo de su propio velo-de-ideas basado en la idea de que «nada puede ser presente a la mente si no es una imagen o percepción» (David Hume, *Philosophical Works* [Boston y Edinburgh], 1854, vol. 4. página 173) y del pirronismo, o escepticismo *excesivo* (pág. 183). Tenía mucho interés en separar el segundo del tercero y en insistir en que no se tomara en serio el escepticismo meramente «profesional» y «técnico» de las «nuevas ideas». Hume no pensaba que estuviera buscando argumentos para apoyar a Sextus; más bien tenía sumo interés en demostrar que el resultado escéptico del pro-

una «teoría del conocimiento» creció en torno a este último problema —el problema de saber si nuestras representaciones internas eran exactas. La idea de una disciplina centrada en «la naturaleza, origen y límites del conocimiento humano» —definición de la «epistemología» tal como aparecía en los manuales— necesitaba un campo de estudio llamado «la mente humana», y ese campo de estudio era lo que había creado Descartes. La mente cartesiana hizo posible simultáneamente el escepticismo del velo-de-ideas y una disciplina consagrada a evitar tal escepticismo.

Con esto no queremos decir que la invención de la mente cartesiana sea condición *suficiente* para el desarrollo de la epistemología. Tal invención nos proporcionó la idea de las representaciones internas, pero esta idea no habría dado lugar a la epistemología sin la confusión que he atribuido a Locke —la confusión, de la que Descartes era en gran parte inocente, entre una explicación mecánica de las operaciones de nuestra mente y la «fundamentación» de nuestras pretensiones de conocimiento. Esto es lo que T. H. Green llamaría

> la confusión fundamental, en que se apoya toda la psicología empírica, entre dos cuestiones esencialmente distintas —una metafísica, cuál es el elemento más simple del conocimiento, y la otra fisiológica, cuáles son las condiciones del organismo humano individual en virtud de las cuales se convierte en vehículo de conocimiento [15].

La distinción de Green entre un «elemento de conocimiento» y «las condiciones del organismo» nos recuerda que una pretensión de conocimiento es la pretensión de haber justificado la creencia, y casi nunca ocurre que recurramos como *justificación* al funcionamiento adecuado de nuestro organismo. Si es cierto que algunas veces justificamos una creencia diciendo, por ejemplo, «Tengo buena vista», ¿por qué vamos a pensar que las «realaciones entre ideas» cronológicas o composicionales, consideradas como *eventos* del espacio interior, podrían decirnos algo sobre las relaciones lógicas entre *proposiciones?* Después de todo, como dice Sellars:

---

yecto de Locke lo que demostraba no era (como creerían Kant y Russell) la necesidad de una nueva y mejor clase de epistemología, sino la necesidad de valorar la poca importancia de la epistemología y la importancia del sentimiento. Sobre el pirronismo en el periodo anterior a Descartes, véase Richard Popkin, *History of Scepticism from Erasmus to Descartes* (Nueva York, 1964).

[15] T. H. Green, Hume and Locke [Green, «Introductions» al *Treatise* de Hume], ed. Ramon Lemon (Nueva York, 1968), pág. 19.

> Al caracterizar un episodio o estado como de *conocimiento,* no estamos presentando una descripción empírica de ese episodio o estado; lo estamos colocando en el espacio lógico de las razones, de justificar y ser capaz de justificar lo que se dice [16].

¿Cómo fue posible que Locke cometiera lo que Sellars llama «un error semejante a la llamada 'falacia naturalista' en ética», el intento de «analizar los hechos epistémicos sin excepción como si fueran hechos no-epistémicos»? [17] ¿Por qué llegó a pensar que una explicación causal de cómo se llega a tener una creencia debía ser índice de la justificación que se tiene de esa creencia?

La respuesta, en mi opinión, es que Locke, y los escritores del siglo XVII en general, *no* creían que el conocimiento fuera una creencia verdadera justificada. Esto se debía a que no pensaban en el conocimiento como si fuera una relación entre una persona y una proposición. Para *nosotros* resulta natural pensar en «lo que sabe S» como reunión de proposiciones que completan las afirmaciones verdaderas de S que comiencen «Yo sé que...». Cuando nos damos cuenta de que los puntos suspensivos se pueden rellenar con material tan variado como «esto es rojo», «$e = mc^2$», «mi Redentor vive» y «me voy a casar con Jane», tenemos razones justificadas para ser escépticos ante la idea de «la naturaleza, origen y límites del conocimiento humano», y de un «departamento de pensamiento» destinado a estudiar este tema. Pero Locke no pensó que el «conocimiento de que» fuera la forma primaria de conocimiento. Creía, como Aristóteles, que el «conocimiento de» era anterior al «conocimiento de que», y, por lo tanto, que el conocimiento era una relación entre personas y objetos más que entre personas y proposiciones. Dada esa imagen, tiene sentido la idea de un examen de nuestra «facultad de entender», como lo tiene la idea de que está capacitada para ocuparse de algunas clases de objetos y no de otras. Todavía tiene más sentido si se está convencido de que esta facultad se parece algo a una tablilla de cera sobre la que los objetos dejan *impresiones,* y si se piensa que «tener una impresión» es en sí mismo *conocer* más que tener un antecedente causal del conocer.

Reid —el gran enemigo en el siglo XVIII de la «idea 'idea'»— se aferró precisamente a la idea de la «impresión», en lo que se-

[16] Wilfrid Sellars, *Science, Perception and Reality* (Londres y Nueva York, 1963), pág. 169.

[17] *Ibíd.,* pág. 131. Véase un desarrollo de la opinión de Sellars, con aplicaciones a los recientes fenomenalismos, en Michael Williams, *Groundless Belief* (Oxford, 1977), especialmente el capítulo segundo.

guiría Green un siglo más tarde, y muchos otros (H. A. Prichard, Wilfrid Sellars, J. L. Austin, Jonathan Bennett) en nuestro siglo XX. Reid dice:

> No hay prejuicio más natural al hombre que pensar que la mente tiene en sus operaciones cierta semejanza con el cuerpo. Por eso, los hombres se han inclinado a imaginar que, de la misma manera que los cuerpos se ponen en movimiento por obra de algún impulso o impresión hechos sobre ellos por cuerpos contiguos, también la mente está hecha para pensar y percibir por alguna impresión realizada sobre ella, o algún impulso ejercido en ella por cuerpos contiguos [18].

Green, en un pasaje que es continuación del párrafo suyo citado más arriba, dice que sólo mediante la confusión entre elementos del conocimiento (proposiciones) y condiciones fisiológicas es posible «describir de alguna manera una idea como 'impresión'... Una metáfora, interpretada como si fuera un hecho, se convierte en la base del sistema filosófico (de Locke)» [19]. Sellars (hablando de Hume más que de Locke) detecta una confusión entre:

1. Una impresión de un triángulo en cuanto ítem rojo y triangular del que se sabe, inmediata y no inferencialmente, que existe y es rojo y triangular

y

2. Una impresión de un triángulo en cuanto conocimiento de que existe un ítem rojo y triangular [20].

Las tres críticas son protestas contra la idea de que una explicación cuasi-mecánica de la forma en que se ven abolladas por el mundo material nuestras tablillas inmateriales pueda ayudarnos a conocer lo que tenemos derecho a creer.

Probablemente, Locke se creía justificado al unir los dos sentidos de «impresión» distinguidos por Sellars, pues creía que las abolladuras de nuestras cuasi-tablillas (como dice Ryle) se autoanunciaban. Así, escribe: «... la impresión, si es que significa algo, no es otra cosa que hacer ciertas las Verdades que se van a percibir. Imprimir algo en la Mente sin que la Mente lo perciba, me parece

---

[18] Thomas Reid, *Essays on the Intellectual Powers of Man,* reimpreso con una introducción de Baruch Brody (Cambridge, Mass., 1969), pág. 100.

[19] Green, *Hume and Locke,* pág. 11.

[20] Wilfrid Sellars, *Philosophical Perspectives* (Springfield, Ill., 1967), pág. 211. página 211.

difícilmente inteligible»[21]. Es como si la *tabula rasa* estuviera perpetuamente bajo el implacable Ojo de la Mente —ya que, como decía Descartes, nada estaba más cerca de la mente que ella misma. Sin embargo, si el sentido de la metáfora *es* éste, resulta evidente que la impresión tiene menos interés que la observación de la marca dejada —todo el conocimiento corre a cargo, por así decirlo, del Ojo que observa la tablilla impresa, más que de la tablilla misma. Por consiguiente, el éxito de Locke dependía de que *no* se desarrollara el sentido de la metáfora, de que se dejara intacta la ambigüedad entre el cuasi-objeto cuasi-rojo-y-triangular del espacio interior y el conocimiento de que dicho objeto estaba ahí. Mientras Aristóteles no había tenido que preocuparse por un Ojo de la Mente, creyendo que el conocimiento era la *identidad* de la mente con el objeto conocido, Locke no disponía de esta alternativa. Como para él las impresiones eran *representaciones,* necesitaba una facultad que fuera *consciente* de las representaciones, una facultad que *juzgara* las representaciones y no simplemente las *tuviera* —juzgara que existían, o que eran fiables, o que tenían tales-y-tales relaciones con otras representaciones. Pero no tenía espacio para ello, pues postular esta facultad habría significado introducir un fantasma en la cuasi-máquina cuyo funcionamiento tenía la esperanza de describir. Mantuvo de Aristótóeles sólo lo necesario para conservar la idea del conocimiento como consistente en que algo parecido a un objeto entraba en el alma[22], pero no lo suficiente para evitar los problemas escépticos sobre la exactitud de las representaciones o las preguntas kantianas sobre la diferencia entre intuiciones con y sin el «pienso». Dicho de otra manera, la conglomerada *mente* cartesiana que Locke daba por supuesta se parecía al νοῦς aristotélico sólo lo suficiente para dar un sabor tradicional a la idea de la «impresión» y se separaba de él lo suficiente para hacer posible el escepticismo de Hume y el trascendentalismo kantiano. Locke estaba haciendo difíciles equilibrios entre el conocimiento-como-identidad-con-el-objeto y el conocimiento-como-juicio-verdadero-sobre-el-objeto, y sólo gracias a esta postura de transición fue posible la confusa idea de la «filosofía moral» en cuanto «ciencia del hombre» empírica[23].

[21] Locke, *Essay,* I, 2, V.

[22] Reid pensaba que Aristóteles, en su doctrina de los fantasmas, había comenzado a bajar por la resbaladiza pendiente que llevaba a Hume. Cfr. *Essays,* página 133.

[23] En su obra pionera *The Emergence of Probability* (Cambridge, 1975), Ian Hacking indica que la idea de *evidencia,* en cuanto relación de confirma-

Otra forma más de describir la tensión existente en el pensamiento de Locke es considerarlo como tendente, por un lado, hacia la fisiología y, por el otro, hacia Aristóteles. Reid y Green, una vez más, coinciden en este diagnóstico. Reid dice de Descartes que

> algunas veces coloca las ideas de los objetos materiales en el cerebro... sin embargo, algunas veces dice que no debemos pensar que las imágenes o huellas del cerebro sean percibidas como si hubiera ojos en el cerebro; estas huellas sólo son ocasiones de que, por las leyes de la unión del alma y el cuerpo, se exciten las ideas en la mente... Descartes parece haber dudado entre las dos opiniones, o haber pasado de una a otra. De forma semejante, parece que Mr. Locke vaciló entre las dos; algunas veces representaba las ideas de las cosas materiales como si estuvieran en el cerebro, pero, con mayor frecuencia, como si estuvieran en la mente misma [24].

Green, examinando las «ideas de la reflexión» de Locke, habla de su «confusión de pensamiento y materia en la imaginaria tablilla cerebral» y dice que

> Locke oculta al dificultad de su vista y de la del lector cambiando continuamente tanto el sujeto receptor como la materia impresora. Comprobamos que la «tablilla» está en continuo retroceso. En primer lugar es la «parte exterior» del órgano corporal. Luego es el cerebro... Luego es la mente perceptiva, que recibe una impresión de la sensación o tiene una idea de la misma. Finalmente, es la mente reflexiva... [25].

La razón de la confusión que critican Reid y Green es que si (como hicieran Aristóteles y Locke) se intenta reproducir en todo el conocimiento el modelo de la percepción sensorial, se produce una división entre la forma literal en que una parte del cuerpo (por ejemplo, la retina) puede tener la misma cualidad que un objeto externo y la forma metafórica en que la persona en su conjunto tiene, por ejemplo, la raneidad «en la mente» si tiene opiniones

---

ción que se da entre proposiciones, sólo estaba comenzando a aparecer en el siglo XVII. Si Hacking está en lo cierto, este hecho arroja bastante luz sobre las distintas y pesadas mezclas de conocimiento proposicional y (supuesto) conocimiento no proposicional en la obra de los empiristas ingleses. Véanse especialmente las observaciones de Hacking sobre Locke, Berkeley y Hume en el capítulo 19.

[24] Reid, *Essays,* págs. 147-148.

[25] Green, *Hume and Locke,* pág. 11; cf. también pág. 163.

sobre las ranas. La idea de una «tablilla inmaterial» descompone la diferencia entre el simple hecho fisiológico y la metáfora especulativa, y toda filosofía que la utilice se verá dividida en ambas direcciones. Es precisamente la elección de la percepción sensorial como modelo, y en especial de las imágenes oculares, lo que hace que Aristóteles y Locke intenten reducir el «conocimiento de que» —creencia justificada y verdadera en forma de proposiciones— a «conocimiento de» interpretado como «tener en la mente». Como Locke se considera a sí mismo un científico de última hora, le encantaría presentar la metáfora de la «tablilla» en términos fisiológicos. Como no puede, su única opción es la mezcla. Cuando se vuelve hacia Aristóteles, comienza a hablar sobre una «mente reflexiva» que es, realmente, muy poco parecida a una tablilla.

Sin embargo, la mezcla más importante en el tratamiento del conocimiento por Locke no es entre cerebro y νοῦς sino, como he dicho, entre el conocimiento y algo que, por ser el simple hecho de *tener* una idea, puede tener lugar sin juicio, y el conocimiento como aquello que resulta de formar juicios justificados. Esta es la mezcla que Kant aborrecía por considerarla el error básico del empirismo —el error más vigorosamente expresado en su crítica de la confusión de «una sucesión de aprehensiones con una aprehensión de la sucesión», pero que hace referencia igualmente a la confusión entre el mero tener dos ideas «yuxtapuestas» —raneidad y verdor— y «sintetizarlas» en el juicio «Las ranas son generalmente verdes». Lo mismo que Aristóteles no tiene un procedimiento claro para relacionar la comprensión de los universales con la emisión de juicios, ni forma ninguna de relacionar la receptividad de las formas en la mente con la construcción de las proposiciones, tampoco Locke tiene nada semejante. Este es el principal defecto de *cualquier* intento de reducir el «conocimiento de que» a «conocimiento de», de considerar el conocimiento con el modelo de la visión.

Este defecto de la doctrina de Locke fue reparado, según Green y la mayoría de los que escribieron de epistemología en el siglo XIX, por Kant. Green resume su propia crítica al empirismo inglés en las siguientes líneas:

> El empirismo de Locke se vuelve invencible en el momento en que se admite que las cosas cualificadas se «encuentran en la naturaleza» sin ninguna acción constitutiva de la mente. Como la única forma efectiva de tratar con Locke es preguntando —Después de la abstracción de todo lo que él mismo admitía que era creación del pensamiento, ¿qué queda que se pueda encontrar sin más?—, Hume debe encontrarse *in limine* con

> la pregunta de si, aparte de aquellas ideas de la relación que, según su propia demostración, no son simples impresiones, es posible al menos la proposición singular. Si no, la singularidad de tal proposición no consiste en ninguna fidelidad de la presentación al sentido[26].

La expresión «acción constitutiva de la mente» es la información de la opinión del propio Green al respecto, que se resume en el eslogan de los idealistas ingleses: Sólo el pensamiento se relaciona. Consideraban esta doctrina como un resumen del eslogan de Kant: «las intuiciones sin los conceptos son ciegas». Se suponía que el descubrimiento de Kant había sido que no hay «cosas cualificadas» —objetos— antes de «la acción constitutiva de la mente». Así pues, un objeto —algo de lo que hay varios predicados verdaderos— es siempre resultado de una síntesis.

Con Kant, el intento de formular una «teoría del conocimiento» recorrió la mitad del camino hacia una concepción del conocimiento como si fuera fundamental «conocimiento de que» y no «conocimiento de» —a medio camino de una concepción del conocimiento que *no* tuviera como modelo a la percepción. Pero, por desgracia, la forma en que Kant realizó el cambio permanecía todavía dentro del marco de referencia cartesiano; se formulaba todavía como una respuesta a la cuestión de cómo podríamos pasar del espacio interior al espacio exterior. Su respuesta paradójica era que el espacio exterior se construía a partir de las *Vorstellungen* que habitaban el espacio interior. Los intentos del siglo XIX por mantener la idea de que el conocimiento es una relación con proposiciones y no con objetos al mismo tiempo que se evitaba la paradoja kantiana, se mantuvieron también dentro del marco de referencia cartesiano, y, por lo tanto, fueron «idealismos». El único objeto que se les podía ocurrir era el constituido por la síntesis de las ideas lockeanas, y por eso mismo se dedicaron a identificar el conjunto de tales objetos con la cosa-en-sí. Por consiguiente, para entender la idea de «epistemología» tal como la heredó el siglo XX, debemos pasar de la confusión de Locke entre explicación y justificación a la confusión de Kant entre predicación (decir algo de un objeto) y síntesis (unir representaciones en el espacio interior).

---

[26] *Ibíd.*, págs. 185-186.

3. La confusión de Kant entre predicación y síntesis

Que una persona forme un juicio predicativo es que llegue a creer que una oración es verdadera. Que un ego transcendental kantiano llegue a creer que una oración es verdadera es que relacione representaciones *(Vorstellungen)* entre sí: dos formas de representaciones radicalmente distintas, conceptos por una parte e intuiciones por la otra. Kant proporcionó un marco de referencia para entender la confusa escena intelectual del siglo XVII cuando dijo que «Leibniz intelectualizó las apariencias, lo mismo que Locke... sensualizó todos los conceptos de la comprensión» [27]. De esta manera creaba la versión clásica de «la historia de la filosofía moderna» según la cual la filosofía pre-kantiana fue una lucha entre el «racionalismo», que quería reducir las sensaciones a conceptos, y el «empirismo», que quería realizar la reducción inversa. Si en vez de eso Kant hubiera dicho que los racionalistas querían dar con una forma de reemplazar las proposiciones sobre cualidades secundarias con proposiciones que realizaran de alguna manera la misma función pero que fueran conocidas con certeza, y que los empiristas se oponían a este proyecto, los dos siguientes siglos de pensamiento filosófico podrían haber sido muy diferentes. Si el «problema del conocimiento» se hubiera formulado en términos de las relaciones entre proposiciones y el grado de certeza que se les atribuía, y no en términos de supuestos *componentes* de las proposiciones, quizá no hubiéramos heredado nuestra idea actual de «historia de la filosofía». Según la historiografía neokantiana clásica, desde la época de Fedón y la Metafísica Z pasando por Abelardo y Anselmo, Locke y Leibniz, y hasta Quine y Strawson la reflexión peculiarmente *filosófica* se ha ocupado de la relación entre universales y particulares. Sin este tema unificador, quizá no habríamos podido ver una problemática continua, descubierta por los griegos y objeto de preocupación hasta nuestros días, y, por lo tanto, quizá no habríamos tenido nunca la idea de la «filosofía» como algo que tiene una historia de dos mil quinientos años. El pensamiento griego y el pensamiento del siglo XVII pueden considerarse tan distintos entre sí y de nuestros intereses como, por ejemplo, la teología hindú y la numerología maya.

No obstante, para bien o para mal, Kant no dio este viraje pragmático. Habló de las representaciones internas más que de las ora-

[27] *K. d. r. V.*, A271, B327.

ciones. Simultáneamente, nos presentó una historia de nuestro tema, fijó su problemática y la profesionalizó (aunque sólo fuera porque hizo imposible que nadie fuera considerado seriamente como «filósofo» sin haber conseguido dominar la primera *Crítica).* Lo hizo introduciendo en nuestra concepción de una «teoría del conocimiento» (y, por tanto, en nuestra concepción de lo que distinguía a los filósofos de los científicos) lo que C. I. Lewis llamaba «una de las ideas filosóficas más antiguas y más universales», a saber:

> Hay, en nuestra experiencia cognitiva, dos elementos; los datos inmediatos, como los de los sentidos, que son presentados o dados a la mente, y una forma, construcción o interpretación, que representan la actividad del pensamiento [38].

Pero esta «idea» no es ni antigua ni universal. No es más antigua que la idea de que poseemos una cosa llamada «experiencia cognitiva». El término *experiencia* ha llegado a convertirse en el término con que los epistemólogos designan un objeto material, un nombre para el conjunto de *cogitationes* cartesianas e ideas lockeanas. En este sentido, «experiencia» es un término del arte filosófico (muy distinto del uso cotidiano, como en «experiencia laboral», en que es equivalente a ἐμπειρία). La afirmación de Lewis de que cuando observamos este conjunto vemos que se divide en dos clases hace que parezca como si al hombre de la calle, sin ayuda de la filosofía, se le pudiera simplemente decir que dirigiera su ojo mental hacia adentro y observara la distinción. Pero un hombre que no sepa que Locke utilizó el término «experiencia» de forma que incluyera únicamente «ideas de sensación y reflexión» y excluyera los juicios, y que Kant lo utilizó para referirse «al objeto y método de conocimiento, la combinación, de acuerdo con las leyes del pensamiento, de todas las funciones del conocimiento» [39], no tendrá claro qué es lo que se supone que deben mirar, y mucho menos cuál es la distinción que debe observar.

Strawson repite la afirmación de Lewis cuando dice que «la dualidad de conceptos generales ...e instancias particulares de conceptos generales encontrados en la experiencia» es una «dualidad

[38] Lewis, *Mind and the World-Order* (Nueva York, 1956), pág. 38.

[39] Heinrich Ratke, *Systematisches Handlexikon zu Kants Kritik der Reinen Vernunft* (Hamburgo, 1929), pág. 62: «Erkenntnis, den denkgestzlichen Zusammenhang aller Funktionen der Erkenntnis». El uso de Kant es tal que ninguna definición menos confusa y menos ricamente ambigua que la de Ratke podría hacerle justicia. Sobre el sentido filosófico de «experiencia», véase John Dewey, *Experience and Nature* (Nueva York, 1958), pág. 11.

fundamental, inevitable en cualquier pensamiento filosófico sobre la experiencia o el conocimiento empírico»[30]. Esta versión es menos engañosa que la de Lewis, sencillamente porque incluye la palabra *filosófico.* La razón de que esta dualidad resulte inevitable en el pensamiento *filosófico* sobre la experiencia es precisamente que los que no dan con ella no se llaman «filósofos». Sólo podemos explicar qué es el «pensamiento filosófico sobre la experiencia» haciendo referencia al mismo tipo de cosas que Kant. Los psicólogos pueden continuar con los estímulos y respuestas, pero esto es *blosse Naturalismus,* y no figura como «filosófico». El sentido común puede hablar de la experiencia en su estilo superficial —examinando si hemos tenido suficiente experiencia de algo para emitir un juicio al respecto, por ejemplo, pero tampoco esto es filosófico. El pensamiento sólo es *filosófico* si, al igual que el de Kant, busca las causas, y no meramente las razones, de las pretensiones de conocimiento empírico, y si la explicación causal resultante es compatible con todo lo que pueda proponer la investigación psicológica[31]. Del pensamiento filosófico del tipo que encuentra inevitable esta dualidad se supone que hace algo más que limitarse a decirnos que normalmente tenemos conocimiento cuando hemos justificado la creencia verdadera, remitiéndonos al sentido común y a la práctica ordinaria para ver los detalles de lo que figura como justificación. Se supone que explica *cómo es posible el conocimiento,* y que lo hace por algún procedimiento a priori que de alguna manera va más allá del sentido común y, sin embargo, evita la necesidad de complicarse con neuronas, o ratas o cuestionarios.

Dados estos requisitos algo exiguos y la ausencia total de conocimiento de la historia de la filosofía, podríamos estar perplejos ante qué era precisamente lo que se quería y por dónde empezar. Esta perplejidad sólo se puede mitigar comprendiendo términos como «Ser frente a Devenir», «sentido frente a intelecto», «percepciones claras frente a confusas», «ideas simples frente a complejas», «ideas e impresiones», «conceptos e intuiciones». Cuando comenza-

---

[30] P. F. Strawson, *The Bounds of Sense* (Londres, 1966), pág. 20.

[31] Puede parecer extraño llamar «causal» a la explicación de Kant, pero la idea de «constitución transcendental» depende totalmente de la idea de Descartes-Locke de la mecánica del espacio interior, y no deberíamos dejarnos engañar por el uso engañoso de «base» en vez de «causa» por parte de Kant. Si eliminamos de Kant lo que Strawson llama «sujeto mítido de la psicología transcendental», no veremos ningún sentido en la revolución copernicana. He estudiado la versión de Strawon sobre lo que queda de Kant cuando se olvida la revolución copernicana en «Strawson's Objectivity Argument», *Review of Metaphysics,* 24 (1970), 207-244.

mos nuestras meditaciones filosóficas, no topamos inevitablemente, como sugieren Lewis y Strawson, con la distinción intuición-concepto. Por supuesto, no sabríamos lo que figuraba como «experiencia», y mucho menos como ser *filosófico* al respecto, si no hubiéramos dominado esa distinción. Como hemos dicho ya, la idea de una «teoría del conocimiento» no tendrá sentido a no ser que hayamos confundido causalidad y justificación al estilo de Locke, e incluso entonces parecerá borrosa mientras no hayamos aislado algunas entidades del espacio interior cuyas relaciones causales parezcan misteriosas. Las entidades requeridas son precisamente los «conceptos» y las «intuiciones». Si Kant hubiera pasado directamente de la idea de que no se debe identificar «la proposición singular» con «la singularidad de una presentación al sentido» (ni, por lo tanto, al intelecto) a una concepción del conocimiento como relación entre personas y proposiciones, no habría necesitado la idea de «síntesis». Podría haber considerado a la persona como una caja negra que emitía oraciones, estando la justificación de estas emisiones en su relación con su entorno (incluyendo las emisiones de sus iguales, las otras cajas negras). Entonces, la pregunta «¿Cómo es posible el conocimiento?» habría sido semejante a la pregunta «¿Cómo son posibles los teléfonos?», en un sentido parecido a «¿Cómo se puede construir algo que haga eso?». En ese caso, la psicología fisiológica, y no la «epistemología», habría sido la única consecuencia legítima de *De Anima* y del *Essay Concerning Human Understanding*.

No obstante, antes de dejar a Kant, es importante preguntar *cómo* consiguió hacer que la distinción concepto-intuición pareciera al mismo tiempo plausible y misteriosamente problemática. Para entender esto, debemos tener en cuenta que la «síntesis» kantiana requerida para un juicio se distingue de la «asociación de ideas» de Hume en que es una relación que sólo puede darse entre ideas de dos clases diferentes —ideas generales e ideas particulares. Así pues, las ideas de «síntesis» y de distinción concepto-intuición están hechas a la medida la una de la otra, habiéndose inventado ambas para dar sentido a la suposición paradójica pero incuestionada que recorre toda la primera *Crítica* —la suposición de que la diversidad es «dada» y la unidad se hace. Dicha suposición se explicita en la afirmación de que el espacio interior contiene algo parecido a lo que Hume encontró en él, un conjunto de «presentaciones singulares al sentido», pero que estas «intuiciones» no se pueden hacer «aparecer a la conciencia» si no son sintetizadas» por un segundo conjunto de representaciones (en que no reparó Hume) —los conceptos— que entran en relaciones uno-muchos con series de intui-

ciones. La razón, no oficial desde luego, de esta suposición es que la exige la estrategia de la revolución copernicana para asegurar que los objetos se conformen a nuestro conocimiento, no que sean capaces de exigir nuestra conformidad [32]. Pero, oficialmente, se utiliza como premisa en la «Deducción Transcendental» para demostrar que la estrategia «copernicana» tiene resultados positivos. Se supone que la «Deducción» demuestra que sólo podemos ser conscientes de los objetos constituidos por nuestra propia actividad sintetizadora. Oficialmente se supone, por tanto, que podemos *ver* que

> de todas las representaciones, la *combinación* es la única que no puede ser dada a través de los objetos... Donde la comprensión no se ha combinado previamente, no se puede descomponer, pues sólo puede ser dado a la facultad de representación algo que admite análisis en cuanto que ha sido combinado por la *comprensión* [33].

Pero, si no hemos leído a Locke y a Hume, ¿cómo sabemos que a la mente se le presenta la diversidad? ¿Por qué hemos de pensar que la sensibilidad, «en su receptividad original» [34], nos ofrece una diversidad, una diversidad que «no puede ser representada como diversidad» [35] hasta que la comprensión haya utilizado conceptos para sintetizarla? No podemos recurrir a la introspección y ver que es así, pues nunca somos conscientes de las intuiciones no sintetizadas, ni de los conceptos independientemente de su aplicación a las intuiciones. La doctrina de que no somos tan conscientes es precisamente el avance de Kant hacia la consideración de que el conocimiento es de proposiciones más que de objetos —lo que le aleja de los intentos de Aristóteles y Locke por reproducir el modelo de la percepción en el conocimiento. Pero si no es un hecho preanalítico evidente que exista tal diversidad, ¿cómo podemos utilizar como premisa la afirmación de que la sensibilidad nos ofrece una diversidad? En otras palabras, ¿cómo sabemos que una diversidad que no se puede representar como diversidad *es* una diversidad?

---

[32] *K. d. r. V.*, Bxvii.

[33] *Ibíd.*, B130.

[34] *Ibíd.*, A100.

[35] *Ibíd.*, A99. Según la *Estética Transcental,* la diversidad es espacio-temporal desde el primer momento. Pero la *Analítica* va contra esto (cfr., por ejemplo, A102, B160n), y el argumento de las *Analogías* no tendría valor si no se abandonara la doctrina de la *Estética.* Cfr. Robert Paul Wolff, *Kant's Theory of Mental Activity* (Cambridge, Mass., 1963), págs. 151 y ss., donde se estudia esta cuestión.

En términos más generales, si queremos demostrar que sólo podemos ser conscientes de las intuiciones sintetizadas, ¿cómo obtenemos nuestra información sobre las intuiciones antes de la síntesis? ¿Cómo sabemos, por ejemplo, que hay más de una? [36]

Se puede responder a esta última pregunta diciendo que si sólo hubiera una la síntesis sería innecesaria. Pero esto nos hace dar vueltas en torno a un pequeño círculo. Lo que queremos saber es si los conceptos *son* sintetizadores, y no sirve de nada que nos digan que no podrían serlo si no hubiera muchas intuiciones que esperan la síntesis. En este punto, creo yo, hemos de confesar que «intuición» y «concepto», en su sentido kantiano, sólo son susceptibles de definiciones contextuales; igual que «electrón» y «protón», sólo tienen sentido como elementos de una teoría que aspira a explicar algo. Pero si admitimos esto, rompemos definitivamente con las llamadas de Locke y Descartes a esa certeza especial con la que somos conscientes de «lo que está más cerca de nuestras mentes» y «más fácil nos es conocer». La suposición de que la diversidad se encuentra y la unidad se hace resulta que tiene como única justificación la afirmación de que sólo esta teoría «copernicana» explicará nuestra capacidad de tener conocimiento sintético a priori [37].

Pero si estimamos que toda la explicación kantiana sobre la síntesis es *solamente* algo que se postula para explicar la posibilidad del conocimiento sintético a priori, si aceptamos la afirmación de que los tejemanejes cuasi-psicológicos descritos en la «Deducción» no tienen base introspectiva, dejará de tentarnos la estrategia «copernicana». La afirmación de que el conocimiento de las verdades necesarias sobre objetos hechos («constituidos») es más inteligible que en el caso de los objetos que se encuentran depende de la suposición cartesiana de que tenemos acceso privilegiado a la actividad de hacer. Pero según la interpretación de Kant que acabamos de ofrecer, no existe tal acceso para nuestras actividades de

[36] Supongamos que un místico nos dice que la intuición nos ofrece la unidad —el blanco resplandor de la eternidad—, mientras que el pensamiento conceptual (como una cúpula de vidrio multicolor) la fragmenta en una multiplicidad. ¿Cómo podríamos saber si era él o Kant quien tenía la razón al hablar de la correlación de la unidad con la receptividad o con la espontaneidad? ¿Importaría mucho?

[37] Dicho de otra manera, el idealismo transcendental del propio Kant echa por tierra la «Deducción Transcendental» —pues o la maquinaria (síntesis) y la materia prima (conceptos, intuiciones) descritas en la «Deducción» son nouménicas, o son fenoménicas. Si son fenoménicas, en contra de lo que dicen las premisas de la «Deducción», podemos ser conscientes de ellas. Si son nouménicas, no se puede saber nada de ellas (incluyendo lo que dice la «Deducción»).

constitución. Sigue vigente el misterio que rodea a nuestro conocimiento de las verdades necesarias. Las entidades teóricas postuladas en el espacio interior no son, por ser interiores, más útiles para explicar cómo puede ocurrir este conocimiento que esas mismas entidades pero del espacio exterior.

4. El conocimiento en cuanto necesitado de «fundamentos»

A esta forma de tratar a Kant se le puede objetar que existe, de hecho, una distinción pre-analítica entre intuiciones y conceptos, tan antigua como Platón. Las intuiciones sensoriales, se podría argumentar, se identifican en primer lugar en cuanto fuente de conocimiento de verdades contingentes, y los conceptos como fuente del conocimiento de verdades necesarias. El conflicto entre racionalismo y empirismo, según esta concepción, no es, como vengo diciendo yo, la forma en que Kant desfiguró a sus predecesores en términos de su nueva distinción, sino que es tan antiguo como el descubrimiento de la enorme diferencia entre verdad matemática y otras verdades más vulgares. He estado hablando como si las conocidas oposiciones entre sentido e intelecto, ideas claras y confusas, etc., fueran parte de un moderno artefacto llamado «teoría del conocimiento». Pero, aunque se reconozca que el sentido «filosófico» de la «experiencia» sea un artefacto moderno, ¿es indudable que la distinción griega entre sentido e intelecto fue un auténtico descubrimiento, tan grande como el de la demostrabilidad rigurosa de la verdad geométrica? ¿Y es cierto que Kant estaba formulando una pregunta interesante cuando preguntaba cómo era posible la verdad necesaria (por ejemplo, matemática)?

Esta objeción me da la oportunidad de introducir un último punto para completar mi explicación del origen y naturaleza de la idea de un «departamento de pensamiento interesado por el origen y naturaleza del conocimiento humano». En mi opinión, Platón no descubrió la distinción entre dos clases de realidad, internas o externas. Más bien, como he indicado anteriormente, fue el primero en articular lo que George Pitcher ha llamado el «Principio Platónico» —que las diferencias en certeza deben corresponder a diferencias en los objetos conocidos [38]. Este principio es una consecuencia natural del intento de considerar el conocimiento siguiendo el modelo de la percepción y de tratar el «conocimiento de» como si

[38] Cfr. capítulo segundo, nota 15.

sirviera de base al «conocimiento de que». Si se supone que necesitamos facultades distintas para «captar» objetos tan diferentes como ladrillos y números (como tenemos distintos órganos sensoriales para los colores y para los olores), el descubrimiento de la geometría parecerá ser el descubrimiento de una nueva facultad llamada νοῦς. A su vez esto engendrará el problema de la razón tal como se expone en el capítulo primero.

El dejarse impresionar por el carácter especial de la verdad matemática forma hasta tal punto parte del «pensar filosóficamente» que es difícil escapar al dominio del Principio Platónico. Sin embargo, si pensamos en la «certeza racional» como la victoria en un argumento más que como la relación con un objeto conocido, al buscar la explicación del fenómeno dirigiremos la vista a nuestros interlocutores más que a nuestras facultades. Si consideramos nuestra certeza sobre el Teorema de Pitágoras como la confianza, basada en la experiencia con argumentos sobre estas materias, de que nadie encontrará una objeción a las premisas de las que lo deducimos, entonces no trataremos de explicarlo por la relación de la razón con la triangularidad. Nuestra certeza será cuestión de conversación entre personas, y no de interacción con la realidad no humana. De esta manera no veremos una diferencia de clase entre verdades «necesarias» y «contingentes». Todo lo más, veremos diferencias en el grado de facilidad con que se pueden poner objeciones a nuestras creencias. En resumen, estaremos donde estaban los sofistas antes de que Platón aplicara su principio e inventara el «pensamiento filosófico»: buscaremos un caso seguro más que un fundamento firme. Estaremos en lo que Sellars llama «el espacio lógico de las razones» más que en el de las relaciones causales con los objetos [30].

En relación con la distinción necesario-contingente, lo que más me interesa destacar es precisamente que la noción de «fundamentos del conocimiento» —verdades que son ciertas debido a sus causas más que debido a los argumentos presentados a su favor— es fruto de la analogía griega (y específicamente platónica) entre percibir y conocer. El rasgo esencial de la analogía es que conocer que una proposición es verdadera debe identificarse con ser causado por un objeto para hacer algo. El objeto a que se refiere la proposición *impone* la verdad de la proposición. La idea de «verdad necesaria»

[30] Véase una exposición favorable a los sofistas y al pensamiento pre-platónico en general, que está de acuerdo con esta concepción, en Laszlo Versényi, *Socratic Humanism* (New Haven, 1963). Véase también lo que dice Heidegger sobre Protágoras en *The Question Concerning Technology and Other Essays,* traducción de William Lovitt (Nueva York, 1977), págs. 143-147.

es precisamente la idea de una proposición que se cree porque el objeto se nos «impone» de forma irresistible. Esta verdad es necesaria en el sentido en que a veces es necesario creer que lo que tenemos delante de los ojos parece rojo —hay una fuerza, no nosotros mismos, que nos compele. Los objetos de las verdades matemáticas no *admiten* juicios ni informes erróneos. Se supone que las verdades tan paradigmáticamente necesarias como los axiomas geométricos no tienen necesidad de justificación de argumentos, de discusión —son tan indiscutibles como el mandato de Zeus cuando desata los rayos, o el de Elena cuando invita a su cama. (Se supone que la ἀνάγκη racional es, por así decirlo, una forma subliminar de la βία bruta.)

Si se prefiere, la idea de «concepto» puede entenderse como «la fuente de conocimiento de las verdades necesarias», pero esto no significa que Lewis y Strawson tengan razón al pensar que la distinción concepto-intuición esté dada preanalíticamente. Es simplemente la versión moderna de un conjunto de metáforas opcionales —las elegidas por Platón, las que se han convertido en definitorias del «pensamiento filosófico». La distinción primaria de Platón no era entre dos clases de realidades del espacio interior, dos clases de representaciones internas. Aunque jugó con metáforas del «espacio interior» (como en la imagen de los pájaros de *Teeteto* y en su uso de «en el alma» ἐν τῇ ψυχῇ), y en ocasiones se aproximó a las imágenes cartesianas del Ojo de la Mente que inspecciona varias imágenes interiores —más o menos llamativas—, su pensamiento fue esencialmente «realista». La distinción platónica a que dio origen la verdad matemática era metafísica más que epistemológica —una distinción entre los mundos del Ser y del Devenir. Lo que corresponde a las distinciones metafísicas en la «línea dividida» de *La República VI* son distinciones no entre clases de representaciones internas no proposicionales, sino entre grados de certeza que incumben a las proposiciones. Platón no se centró en la idea de las realidades internas no proposicionales, sino más bien en la de las diversas partes del alma y del cuerpo que se ven compelidas a su manera respectiva por sus respectivos objetos. Platón, como Descartes, basó un modelo del hombre en la distinción entre dos clases de verdades, pero estos dos modelos eran completamente diferentes. Pero lo que más importa es que la idea de que la existencia de la verdad matemática exija semejante modelo explicativo no es algo pre-analítico, dado al comienzo de la reflexión filosófica. Es producto de la elección de un determinado conjunto de metáforas para hablar sobre el

conocimiento, las metáforas basadas en la percepción que subyacen a las discusiones platónicas y modernas [40].

Y ya hemos hablado bastante sobre la objeción con que he comenzado la sección. Ahora quiero detenerme en la cuestión de que la idea de «fundamentos del conocimiento» es producto de la elección de metáforas perceptivas. Recapitulemos: podemos pensar en el conocimiento como relación con proposiciones, y por tanto en la justificación en cuanto relación entre las proposiciones en cuestión y otras proposiciones de las que se pueden deducir las primeras. O podemos considerar el conocimiento y la justificación como relaciones privilegiadas con los objetos de que tratan dichas proposiciones. Si pensamos de la primera manera, no veremos ninguna necesidad de terminar el retroceso potencialmente infinito de proposiciones-presentadas-en-defensa-de-otras-proposiciones. Sería absurdo seguir hablando del tema una vez que todos, o la mayoría, o los sabios, estuvieran satisfechos, pero *podemos* hacerlo, por supuesto. Si concebimos el conocimiento de la segunda forma, desearemos pasar de las razones a las causas, del argumento a la compulsión del objeto conocido, hasta llegar a una situación en la que el argumento no sólo fuera absurdo sino imposible, pues el que se sienta atrapado por el objeto en la forma requerida será *incapaz* de dudar o de ver una alternativa. Llegar a ese punto es llegar a los fundamentos del conocimiento. Para Platón, se llegaba a ese punto escapando de los sentidos y abriendo de par en par la facultad de la razón —el Ojo del Alma— al Mundo del Ser. Para Descartes, era cuestión de apartar el Ojo de la Mente de las representaciones internas confusas y dirigirlo a las claras y distintas. Con Locke, se trataría de invertir las direcciones de Descartes y ver las «presentaciones singulares al sentido» como lo que debería «agarrarnos» «aquello de lo que no podemos ni deseamos escapar». Antes de Locke, a nadie se le habría ocurrido buscar los fundamentos del conocimiento en el mundo de los sentidos. Aristóteles, es cierto, había señalado que no podemos estar equivocados sobre cómo nos parecen las cosas, pero la idea de basar el conocimiento en las apariencias le habría parecido, a él y a Platón, una idea absurda. Lo que queremos tener como objeto de conocimiento es precisamente lo que no es una apariencia, y para ninguno de los dos habría tenido sentido la idea de que las proposiciones sobre una clase de objeto (las apariencias) fueran *pruebas* para las proposiciones sobre un objeto de otra clase (lo que hay realmente).

---

[40] Resulta obvio que mi exposición de las metáforas oculares platónicas debe mucho a Dewey y a Heidegger.

Después de Descartes, cada vez se insistió menos en la distinción apariencia-realidad, que fue sustituida por la distinción interior-exterior. La pregunta «¿Cómo puedo escapar de la esfera de la apariencia?» fue sustituida por esta otra: «¿Cómo puedo escapar de detrás del velo de ideas?» Locke ofreció esta respuesta: Utiliza tu certeza sobre cómo aparecen las cosas a tus sentidos de la misma manera que Platón utilizó los axiomas de geometría —es decir, como premisa para inferir todo lo demás (únicamente en forma inductiva, y no, como en el caso de Platón, deductivamente). Esta respuesta sólo pareció adecuada mientras Hume no se puso a trabajar con ella, pero tenía cierto encanto inocente. Satisfacía la misma necesidad de sentirse atrapado, dominado y compelido que había experimentado Platón, pero, sin embargo, «las ideas simples de la sensación» parecían menos pretenciosas y más puestas al día que las Formas platónicas. Por consiguiente, en la época de Kant parecía como si hubiera dos fundamentos posibles del conocimiento —había que elegir entre la versión interiorizada de las Formas, las ideas claras y distintas, por una parte, y las «impresiones» de Hume, por la otra. En ambos casos, se elegían objetos para sentirse dominados por ellos. Kant, al rechazar ambos supuestos objetos como esencialmente incompletos e incapaces para dominar a no ser que se combinaran mutuamente en una «síntesis», fue el primero en concebir los fundamentos del conocimiento como proposiciones en vez de como objetos. Antes de Kant, una investigación de «la naturaleza y origen del conocimiento» había sido una búsqueda de representaciones internas privilegiadas. Con Kant, se convirtió en la búsqueda de las reglas que la mente se había impuesto (los «Principios de la Comprensión Pura»). Esta es una de las razones por las que se pensó que Kant nos había llevado de la naturaleza a la libertad. En vez de vernos como máquinas cuasi-newtonianas, a la espera de ser compelidas por las realidades internas adecuadas y, por lo tanto, de funcionar en conformidad al designio que la naturaleza tenía para nosotros, Kant hizo posible que nos viéramos en actitud de decidir (numeralmente, y, por tanto, inconscientemente) cómo había que dejar que fuera la naturaleza.

Sin embargo, Kant no nos libró de la confusión de Locke entre justificación y explicación causal, confusión básica contenida en la idea de «teoría del conocimiento». La idea de que nuestra libertad depende de una epistemología idealista —de que para vernos «subir por encima de los mecanismos» debemos ponernos en plan transcendental y afirmar que somos nosotros mismos quienes hemos «constituido los átomos y el vacío»— es precisamente una vuelta al error

de Locke. Es suponer que el espacio lógico de presentación de razones —de justificar nuestras elocuciones y nuestras acciones— tiene que estar en alguna relación especial con el espacio lógico de la explicación causal para asegurar o bien un acuerdo entre los dos (Locke) o la incapacidad de uno para interferir en el otro (Kant). Kant tenía razón cuando pensaba que el acuerdo no tenía sentido y la interferencia era imposible, pero se equivocaba al pensar que para establecer el último punto hacía falta la idea de la «constitución» de la naturaleza por el sujeto cognoscente. El progreso de Kant en sentido de una concepción del conocimiento proposicional y no perceptiva se quedó a mitad de camino, pues estaba contenido dentro del marco de referencia de las metáforas causales —«constitución», «hacer», «configurar», «sintetizar», y otras semejantes.

La diferencia entre la línea principal de la tradición anglosajona y la de la tradición alemana en la filosofía del siglo XX es la expresión de dos posturas opuestas hacia Kant. La tradición que se remonta a Russell descalificaba el problema de Kant sobre las verdades sintéticas a priori por considerarlo como una forma equivocada de entender la naturaleza de las matemáticas, y, por lo tanto, consideraba la epistemología como si fuera esencialmente cuestión de poner al día a Locke. En el proceso de esta actualización, la epistemología quedó separada de la psicología, por ser considerada como un estudio de las relaciones evidentes entre proposiciones básicas y no básicas, y estas relaciones se consideraban como una cuestión de «lógica» más que de hechos empíricos. En la tradición alemana, por el contrario, la defensa de la libertad y la espiritualidad mediante la noción de «constitución» siguió en pie como misión distintiva de la filosofía. El empirismo lógico y, más tarde, la filosofía analítica fueron rechazadas por la mayoría de los filósofos alemanes (y muchos franceses) por considerar como no «transcendental», y, por tanto, no sería ni metodológicamente válida ni debidamente edificante. Incluso los que tuvieron dudas muy serias sobre la mayoría de las doctrinas kantianas no dudaron nunca de que era esencial algo parecido a su «giro transcendental». Por parte anglo-sajona, se pensó que el llamado giro lingüístico realizaría la tarea de separar la filosofía de la ciencia, al mismo tiempo que eliminaría cualquier vestigio, o tentación, de «idealismo» (considerado como el pecado dominante de la filosofía del Continente).

De todas las maneras, la mayoría de los filósofos de ambos lados del Canal han seguido siendo kantianos. Aun cuando afirman que han «ido más allá» de la epistemología, han llegado al acuerdo de que la filosofía es una disciplina que se encarga del estudio de los

aspectos «formales» o «estructurales» de nuestras creencias, y que cuando las examina el filósofo realiza la función cultural de mantener la integridad de las demás disciplinas, limitando sus afirmaciones a lo que puede «fundarse» adecuadamente. Las grandes excepciones a este consenso neo-kantiano son, una vez más, Dewey, Wittgenstein y Heidegger. En relación con el tema de esta sección —la idea de «fundamentos» del conocimiento en cuanto basada en una analogía con la compulsión a creer cuando se mira un objeto— resulta especialmente importante Heidegger. Este ha tratado de demostrar cómo la idea de «objetividad» procede, como él dice, de la «identificación [platónica] de φυσις e ἰδέα» —de la realidad de una cosa con su presencia ante nosotros [41]. Le interesa explorar la forma en que el Occidente llegó a obsesionarse con la idea de nuestra relación primaria con los objetos como si fuera análoga con la percepción visual, y, por tanto, señala que podría haber otras concepciones de nuestras relaciones con las cosas. Las raíces históricas de la analogía, de Aristóteles-Locke, del conocimiento con la percepción quedan fuera del ámbito de este libro, pero al menos puedo tomar de Heidegger la idea de que el deseo de una «epistemología» es simplemente el producto más reciente del desarrollo dialéctico de un conjunto de metáforas elegido originalmente [42].

Describir este desarrollo como una secuencia lineal es, ciertamente, simplista, pero quizá sirva para pensar que la metáfora dominante original era la de determinar nuestras creencias poniéndolas cara a cara con el objeto de la creencia (la figura geométrica que prueba el teorema, por ejemplo). El siguiente paso es pensar que entender cómo se conoce mejor es entender cómo se mejora la actividad de una facultad cuasi-visual, el Espejo de la Naturelaza, y, por lo tanto, pensar en el conocimiento en cuanto agrupación de representaciones exactas. Luego viene la idea de que la forma de tener representaciones exactas es encontrar, dentro del Espejo, una clase privilegiada especial de representaciones tan irresistibles que no se puede dudar de su exactitud. Estos fundamentos privilegiados serán los fundamentos del conocimiento, y la disciplina que nos dirija hacia ellos —la teoría del conocimiento— será el fundamento de la cultura. La teoría del conocimiento será la búsqueda de lo que obliga a la mente a creer en el mismo momento en que es desvelado. La filosofía-como-epistemología será la búsqueda de estructu-

---

[41] Martin Heidegger, *Introduction to Metaphysics*, traducción de Ralph Manheim (New Haven, 1959), pág. 185.

[42] Martin Heidegger, *The End of Philosophy*, ed. y trad. de Joan Stambaugh (Nueva York, 1973), pág. 88.

ras inmutables dentro de las cuales deban estar contenidos el conocimiento, la vida y la cultura —estructuras establecidas por las representaciones privilegiadas que estudia. Así pues, el consenso neokantiano parece el producto final de un deseo original de sustituir la *conversación* por la *confrontación* en cuanto determinante de nuestra creencia.

En la parte III trato de demostrar cómo parecen las cosas si se considera suficiente la conversación y se abandona la búsqueda de confrontación, y, por tanto, si *no* se concibe el conocimiento como representaciones en el Espejo de la Naturaleza. No obstante, en los tres siguientes capítulos intentaré en primer lugar (en el capítulo cuarto) esbozar la forma que adoptó el consenso neo-kantiano en la filosofía del siglo XX, y la confusión en que ha caído recientemente una forma de este consenso —la «filosofía analítica»—. Para ello tendremos que describir, y defender, los ataques de Quine y Sellars a la idea de la representación privilegiada. En los capítulos 5 y 6 considero dos posibles sucesores de la filosofía-como-epistemología —la psicología empírica y la filosofía del lenguaje, respectivamente— que se mantienen dentro del consenso neo-kantiano al considerar que la filosofía es, paradigmáticamente, el estudio de la representación.

CAPÍTULO IV

Representaciones privilegiadas

1. VERDAD APODÍCTICA, REPRESENTACIONES PRIVILEGIADAS Y FILOSOFÍA ANALÍTICA

Al acabar el siglo XIX, los filósofos estaban preocupados, y con razón, por el futuro de su disciplina. Por una parte, el nacimiento de la psicología empírica había planteado la cuestión siguiente: «¿Qué necesitamos saber sobre el conocimiento que no pueda decírnoslo la psicología?»[1]. Desde el intento de Descartes de conseguir que el mundo fuera un lugar seguro para las ideas claras y distintas y el de Kant por que lo fuera para las verdades sintéticas a priori, la ontología había sido dominada por la epistemología. Por eso, la «naturalización» de la epistemología por la psicología indicaba que un fisicismo simple y relejado podría ser la única clase de concepción ontológica necesaria. Por otra parte, la tradición del idealismo alemán había venido a parar —en Inglaterra y América— a lo que se ha descrito adecuadamente como «una continuación del Protestantismo por otros medios». Los idealistas se proponían salvar los «valores espirituales» que el fisicismo parecía descuidar, invocando argumentos de Berkeley para liberarse de la sustancia material y argumentos hegelianos para liberarse del ego individual (al mismo tiempo que prescindían totalmente del historicismo de Hegel). Pero

[1] Esta pregunta se ha vuelto a repetir en nuestro siglo bajo formas que describiremos en el siguiente capítulo. La psicología nació de la filosofía con la confusa esperanaz de poder retroceder más allá de Kant y recuperar la inocencia lockeana. Desde entonces, los psicólogos han protestado inútilmente por el menosprecio manifestado por los filósofos neo-kantianos (de corte analítico y fenomenológico).

pocos se tomaron en serio estos nobles esfuerzos. El sincero reduccionismo de Bain y Mill y el igualmente sincero romanticismo de Royce llevó a los ironistas estéticos como James y Bradley, así como a reformadores sociales como el joven Dewey, a proclamar la irrealidad de los problemas y soluciones epistemológicos tradicionales. Se sintieron impulsados a la crítica radical de «la verdad como correspondencia» y del «conocimiento como exactitud de las representaciones», amenazando de esta manera toda la concepción kantiana de la filosofía como metacrítica de las disciplinas particulares. Al mismo tiempo, filósofos tan distintos como Nietzsche, Bergson y Dilthey estaban minando algunas de las presuposiciones kantianas. Por algún tiempo, dio la impresión de que la filosofía podía alejarse de una vez por todas de la epistemología, de la búsqueda de certeza, estructura y rigor, y del intento de constituirse como tribunal de la razón.

Sin embargo, el espíritu festivo que parecía a punto de invadir la filosofía alrededor del año 1900, quedó cortado de raíz. Igual que la matemática había inspirado a Platón a inventar el «pensamiento filosófico», los filósofos más graves volvieron la vista a la lógica matemática para rescatarla de la sátira exuberante de sus críticos. Las figuras paradigmáticas en este intento de recuperar el espíritu matemático fueron Husserl y Russell. Husserl vio a la filosofía atrapada entre el «naturalismo» y el «historicismo», ninguno de los cuales ofrecía la clase de «verdades apodícticas» que Kant había dicho que constituían patrimonio de los filósofos [2]. Russell se unió a Husserl en su denuncia del psicologismo que había infectado la filosofía de las matemáticas, y anunció que la lógica era la esencia de la filosofía [3]. Llevado por la necesidad de encontrar algo sobre

[2] Cfr. Edmund Husserl, «Philosophy as Rigorous Science», en *Phenomenology and the Crisis of Philosophy,* edición y traducción de Quentin Lauer (Nueva York, 1965), pág. 120. En este ensayo, publicado en 1910, Husserl analizaba el naturalismo repitiendo el ataque a las concepciones psicológicas de la lógica presentado en su *Logical Investigations.* (Cfr. págs. 80 y ss. sobre la auto-refutación del naturalismo mediante su reducción de las normas a hechos.)

[3] Bertrand Russell terminaba el capítulo titulado «Logic as the Essence of Philosophy» de su *Our Knowledge of the External World* (Londres, 1914) con estas afirmaciones:

> La antigua lógica encadenaba el pensamiento, mientras que la nueva lógica le da un mayor impulso. En mi opinión, ha introducido en la filosofía el mismo avance que Galileo introdujera en la física, haciendo posible por fin ver qué clases de problemas pueden tener solución, y cuáles deben abandonarse por superar las capacidades humanas. Y donde parece posible una solución, la nueva lógica proporciona un método

lo que ser apodíctico, Russell descubrió la «forma lógica» y Husserl descubrió las «esencias», los aspectos «puramente formales» del mundo que permanecían cuando «se ponían entre paréntesis» los aspectos no formales. El descubrimiento de estas representaciones privilegiadas puso una vez más en marcha una búsqueda de seriedad, pureza y rigor [4], búsqueda que duró unos cuarenta años. Pero, al final, los seguidores heterodoxos de Husserl (Sartre y Heidegger) y los seguidores heterodoxos de Russell (Sellars y Quine) presentaron las mismas preguntas sobre la posibilidad de la verdad apodíctica que Hegel había formulado con relación a Kant. La fenomenología se fue transformando gradualmente en lo que Husserl llamaba, con desesperación, «mera antropología» [5], y la epistemología «analítica» (es decir, la «filosofía de la ciencia») se hizo cada vez más historicista y menos «lógica» (como ocurrió en Hanson, Kuhn, Harré, y Hesse). Así pues, setenta años después de la «Filosofía como Ciencia Rigurosa» de Husserl y de la «Lógica como Esencia de la Filosofía» de Russell, hemos vuelto a los mismos supuestos peligros con que se enfrentaron los autores de estos manifiestos: si la filo-

---

> que nos permite obtener resultados que no expresan meros particularismos personales, sino que deben recibir la aprobación de todos los que están preparados para formarse una opinión.

En el contexto en que estamos, la acusación clásica (presentada, por ejemplo, por Dummett y por Anscombe) de que Russell confundía las doctrinas específicamente semánticas de Frege y Wittgenstein, que *sí* procedían de la nueva lógica, con las doctrinas epistemológicas que no procedían de ella, carece de importancia. La acusación es bastante razonable, pero sin esta confusión el movimiento analítico o no habría conseguido remontar el vuelo o habría sido algo completamente distinto. Sólo en las dos últimas décadas se ha comenzado a trazar una distinción clara entre «filosofía lingüística» y «filosofía del lenguaje». Véase el capítulo sexto, sección I, donde se habla de esta distinción.

[4] Véase Russell, *Our Knowledge of the External World*, pág. 61 (en la edición americana [Nueva York, 1924]), y Husserl, *Phenomenology*, páginas 110-111.

[5] Véase Herbert Spiegelberg, *The Phenomenological Movement*, segunda edición (La Haya, 1965), I, págs. 275-283, y David Carr, «Translator's Introduction» [Introducción del traductor], en Edmund Husserl, *The Crisis of European Sciences and Transcendental Phenomenology* (Evanston, 1970), páginas XXV-XXXVIII. Véase también la reacción de Ryle a *Sein und Seit*, que es un ejemplo del parentesco entre los proyectos anglosajones influenciados por Russell y el proyecto original de Husserl: «Mi opinión personal es que, en cuanto Filosofía Primera, la Fenomenología va actualmente camino de la bancarrota y el desastre y acabará o en un Subjetivismo ruinoso o en un Misticismo ampuloso» *(Mind,* 1929; citado por Spiegelberg, I, pág. 347). La premonición de Ryle era que la llegada de la «fenomenología existencial» significaba el fin de la fenomenología como «ciencia rigurosa».

sofía se vuelve demasiado naturalista, las disciplinas positivas más estrictas la dejarán de lado; si se vuelve demasiado historicista, quedará engullida por la historia intelectual, la crítica literaria y otros lugares borrosos dentro de «las humanidades» [6].

No entra dentro de los objetivos de este libro hacer una historia completa de las glorias y miserias de la fenomenología y de la filosofía analítica. En este capítulo me limitaré a exponer cómo, en los momentos finales del movimiento analítico, se desprestigió la idea de las dos clases de representaciones —intuiciones y conceptos. Lo que he venido afirmando hasta ahora es que la imagen kantiana de que los conceptos e intuiciones se unían para producir el conocimiento es necesaria para dar sentido a la idea de la «teoría del conocimiento» en cuanto disciplina específicamente filosófica, distinta de la psicología. Esto equivale a decir que si no tenemos la distinción entre lo que es «dado» y lo que es «añadido por la mente», o entre lo «contingente» (por estar influenciado por lo que es dado) y lo «necesario» (por estar totalmente «dentro de» la mente y bajo su control), no sabremos qué es lo que puede figurar como «reconstrución racional» de nuestro conocimiento. No sabremos cuál puede ser el objetivo o el método de la epistemología. Estas dos distinciones fueron atacadas, con algunos intervalos, a lo largo de toda la historia del movimiento analítico. Neurath había puesto en duda el recurso de Carnap a lo dado, por ejemplo, y muchas veces se habían expresado dudas sobre la idea de «conocimiento por contacto», de Russell, y la de «lenguaje expresivo», de Lewis. Sin embargo, estas dudas sólo lograron abrirse paso en los primeros años 50,

---

[6] Creo que en Inglaterra y América la filosofía ya ha sido desplazada por la crítica literaria en su principal función cultural, la de ser fuente de la auto-descripción que los jóvenes se hacen de su propia diferencia con el pasado. Cfr. Harold Bloom, *A Map of Misreading* (Nueva York, 1975), pág. 39:

> El profesor de literatura en la América de nuestros días, mucho más que el profesor de historia o filosofía o religión, está condenado a enseñar el carácter presente del pasado, pues la historia, la filosofía y la religión se han retirado de la Escena de la Instrucción en cuanto agentes, dejando al desconcertado profesor de literatura en el altar, preguntándose aterrorizado si él va a ser el sacerdote o el sacrificio.

Esto se debe, en términos generales, al tono kantiano y antihistoricista de la filosofía anglosajona. La función cultural de los profesores de filosofía en países donde no se ha olvidado a Hegel es muy diferente, y más próxima a la posición de los críticos literarios en América. Véase mi «Professionalized Philosophy and Transcendentalist Culture», *Georgia Review*, 30 (1976), páginas 757-769.

con la aparición de *Investigaciones filosóficas* de Wittgenstein, la burla que hizo Austin de «la ontología de la diversidad sensible», y «El empirismo y la filosofía de la mente» de Sellars. La distinción entre lo necesario y lo contingente —revitalizada por Russell y el Círculo de Viena en forma de distinción entre «verdadero en virtud del significado» y «verdadero en virtud de la experiencia»— se había mantenido generalmente al margen de las discusiones, y había formado el mínimo común denominador del análisis del «lenguaje ideal» y el «lenguaje común». Sin embargo, también en los primeros años 50, Quine («Two Dogmas of Empiricism») puso en tela de juicio esta distinción, y con ella la idea comúnmente admitida (entre otros por Kant, Husserl y Russell) de que la filosofía era para la ciencia empírica como el estudio de la estructura para el estudio del contenido. Dadas las dudas de Quine (fortalecidas por otras dudas semejantes formuladas en las *Investigaciones* de Wittgenstein) sobre cómo se sabe cuándo estamos respondiendo a la compulsión del «lenguaje» y no a la de la «experiencia», cada vez fue más difícil explicar en qué sentido la filosofía tenía un «ámbito» formal de investigación, y, por lo mismo, cómo era posible que sus resultados tuvieran el carácter apodíctico deseado. Estos dos desafíos eran desafíos a la misma idea de una «teoría del conocimiento», y, por tanto, a la filosofía misma, concebida como disciplina que se centra en dicha teoría.

En las páginas siguientes, me limitaré a exponer dos formas radicales de criticar los fundamentos kantianos de la filosofía analítica —la crítica conductista hecha por Sellars a «todo el marco de referencia de lo dado» y el enfoque conductista de Quine ante la distinción necesario-contingente. Presentaré a las dos como formas de holismo. Mientras se conciba el conocimiento como una representación precisa —como el Espejo de la Naturaleza—, las doctrinas holísticas de Quine y Sellars parecen absurdamente paradójicas, pues dicha precisión requiere una teoría de las representaciones privilegiadas, que sean automática e intrínsecamente precisas. Por eso, la respuesta a Sellars en el caso de lo dado y a Quine en el de la analicidad suele ser que «han ido demasiado lejos» —que han dejado que el holismo les trastorne y les haya hecho dejar de lado el sentido común. Para defender a Sellars y a Quine, trataré de demostrar que su holismo es producto de su compromiso con la tesis de que la justificación no es cuestión de una relación especial entre ideas (o palabras) y objetos, sino de conversación, de práctica social. La justificación conversacional, por así decirlo, es naturalmente holística, mientras que la idea de justificación incrustada en la tradición

epistemológica es reductiva y atomista. Trataré de hacer ver cómo Sellars y Quine invocan el mismo argumento, un argumento dirigido igualmente contra las distinciones dado-frente-a-no-dado y necesario-frente-a-contingente. La premisa fundamental de este argumento es que entendemos el conocimiento cuando entendemos la justificación social de la creencia, y, por tanto, no tenemos ninguna necesidad de considerarlo como precisión en la representación.

Una vez que la conversación sustituye a la confrontación, se puede descartar la idea de la mente como Espejo de la Naturaleza. Entonces resulta ininteligible la idea de la filosofía como disciplina que busca representaciones privilegiadas entre las que constituyen el Espejo. Un holismo total no tiene lugar para la idea de la filosofía como «conceptual», como «apodíctica», como descubridora de los «fundamentos» del resto del conocimiento, como explicación de cuáles son las representaciones «puramente dadas» o «puramente conceptuales», como presentadora de una «notación canónica» y no de un descubrimiento empírico, o como aislamiento de «categorías heurísticas que están más allá del marco de referencia». Si vemos el conocimiento como una cuestión de conversación y de práctica social, más que como un intento de reflejar la naturaleza, no es probable que pretendamos llegar a una meta-práctica que sea la crítica de todas las formas posibles de práctica social. Así pues, el holismo produce, como ha explicado detalladamente Quine y Sellars ha señalado de pasada, una concepción de la filosofía que no tiene nada que ver con la búsqueda de certeza.

Sin embargo, ni Quine ni Sellars han elaborado una nueva concepción de la filosofía con cierto detalle. Quine, tras afirmar que no hay separación entre ciencia y filosofía, suele suponer que con ello ha demostrado que la ciencia puede reemplazar a la filosofía. Pero no queda claro cuál es la tarea que señala a la ciencia. Y tampoco es claro por qué la ciencia natural, y no las artes, o la política, o la religión, deba ocupar el espacio que queda vacante. Además, su concepción de la ciencia sigue siendo curiosamente instrumentalista. Está basada en una distinción entre «estímulos» y «postulados» que parece servir de ayuda y tranquilidad a la antigua distinción intuición-concepto. Sin embargo, Quine transciende ambas distinciones al admitir que las estimulaciones de los órganos sensoriales tienen tanto de «postulados» como cualquier otra cosa. Es como si Quine, tras haber renunciado a las distinciones conceptual-empírico, analítico-sintético y lenguaje-hecho, no fuera todavía capaz de renunciar totalmente a la que existe entre lo dado y lo postulado. Por el contrario, Sellars, tras haber vencido la última distinción, no pue-

de renunciar del todo al primer grupo. A pesar de que reconoce cortésmente el triunfo de Quine sobre la analicidad, los escritos de Sellars siguen impregnados de la idea de «presentar el análisis» de varios términos u oraciones, y de un uso tácito de la distinción entre lo necesario y lo contingente, lo estructural y lo empírico, lo filosófico y lo científico. Cada uno de estos autores suele utilizar de forma continua, no oficial, tácita y heurística, la distinción que el otro ha transcendido. Es como si no fuera posible escribir de filosofía analítica sin *una* al menos de las dos grandes distinciones kantianas, y como si ni Quine ni Sellars estuvieran dispuestos a romper los últimos vínculos que los unen a Russell, Carnap y la «lógica en cuanto esencia de la filosofía».

Me da la impresión de que la filosofía analítica no se *puede* escribir sin una u otra de estas distinciones. Si no hay intuiciones en que se puedan resolver los conceptos (a la manera del *Aufbau)* ni relaciones internas entre los conceptos que hagan posibles los «descubrimientos gramaticales» (a la manera de la «filosofía de Oxford»), cuesta mucho imaginarse qué puede ser un «análisis». Acertadamente, son ya pocos los filósofos analíticos que tratan de explicar qué es ofrecer un análisis. Aunque en los años 30 y 40, bajo la égida de Russell y Carnap, hubo gran cantidad de publicaciones metafilosóficas, y otra avalancha de publicaciones semejantes en los años 50, cuyos paradigmas fueron *Philosophical Investigations* y *The Concept of Mind*[7], ahora son pocos los intentos de llevar la «filosofía analítica» a la auto-conciencia explicando cómo distinguir un buen análisis de uno malo. La actual ausencia de reflexión metafilosófica dentro del movimiento analítico es, creo yo, sintomática del hecho sociológico de que la filosofía analítica es ahora, en varios países, la escuela de pensamiento más arraigada. Así, en estos países, *todo* lo que hacen filósofos que utilizan un cierto estilo, o mencionan ciertos temas, se interpreta *(ex officiis suis,* por así decirlo) como continuación de la obra iniciada por Russell y Carnap. Una vez que un movimiento radical se impone sobre el *establishment* contra el que se rebeló, hay menos necesidad de autoconciencia metodológica, de auto-crítica, o de un sentido de la localización en el espacio dialéctico o en el tiempo histórico.

No creo que exista ya nada que sea identificable como «filosofía analítica», como no sea de alguna forma estilística o sociológica. Esta observación no tiene sentido despectivo, como si se hubiera

---

[7] He tratado de resumir estas publicaciones, hasta 1965, en la introducción a *The Linguictic Turn,* ed. Richard Rorty (Chicago, 1967).

frustrado una expectación legítima. El movimiento analítico dentro de la filosofía (como cualquier movimiento en cualquier disciplina) agotó las consecuencias dialécticas de un conjunto de suposiciones, y ahora tiene poco más que hacer. La fe optimista que Russell y Carnap compartieron con Kant —que la filosofía, descubierto por fin su método adecuado y su esencia, se había colocado finalmente en el sendero seguro de la ciencia— no es algo que se deba ridiculizar ni deplorar. Este optimismo sólo es posible para hombres de gran imaginación y audacia, los héroes de su época.

2. Conductismo epistemológico

La forma más sencilla de describir los rasgos comunes de los ataques de Quine y Sellars al empirismo lógico es decir que ambos plantean cuestiones conductistas sobre el privilegio epistémico que el empirismo lógico pretende para ciertas afirmaciones, en cuanto informes de representaciones privilegiadas. Quine pregunta cómo ha de discriminar un antropólogo las oraciones a las que asienten invariablemente y de forma incondicional los nativos, distinguiendo los tópicos empíricos, por una parte, y las verdades conceptuales necesarias, por la otra. Sellars pregunta cómo la autoridad de los informes en primera persona de, por ejemplo, cómo nos aparecen las cosas, de los dolores que nos hacen sufrir, y de los pensamientos que vagan por nuestras mentes, difiere de la autoridad de los informes de expertos sobre la fatiga de los metales, la conducta de apareamiento de los pájaros o los colores de los objetos físicos, por ejemplo. Podemos unir ambas preguntas en una: «¿Cómo saben nuestros semejantes cuáles son aquellas de nuestras afirmaciones en que pueden aceptar nuestra palabra y cuáles las que exigen posterior confirmación?» Parecería suficiente que los nativos supieran qué oraciones son incuestionablemente verdaderas, sin saber cuáles son verdaderas «en virtud del lenguaje». Parecería suficiente que nuestros semejantes creyeran que no hay mejor forma de informarse de nuestros estados interiores que a partir de nuestros informes, sin que supieran qué «hay detrás» de nuestra formulación. Parecería también suficiente que *nosotros* supiéramos que nuestros semejantes tienen esta actitud complaciente. Sólo eso parece suficiente para esa certeza interior sobre nuestros estados interiores que la tradición ha explicado mediante «la presencia inmediata a la conciencia», «el sentido de la evidencia», y otras expresiones de la suposición de que los reflejos en el Espejo de la Naturaleza son intrínsecamente

mejor conocidos que la naturaleza misma. Para Sellars, la certeza de «Tengo un dolor» es un reflejo del hecho de que nadie se molesta en ponerlo en duda, no al revés. Lo mismo ocurre, según Quine, con la certeza de «Todos los hombres son animales» y de «Ha habido perros negros». Quine piensa que los «significados» desaparecen como ruedas que no forman parte del mecanismo [8], y Sellars opina lo mismo de los «episodios no-verbales que se auto-autentifican» [9]. Más en general, si las afirmaciones son justificadas por la sociedad más que por el carácter de las representaciones internas que expresan, no tiene sentido tratar de aislar las representaciones *privilegiadas*.

El explicar la racionalidad y autoridad epistemológica por referencia a lo que la sociedad nos permite decir, y no lo segundo por lo primero, constituye la esencia de lo que designaré con el nombre de «conductismo epistemológico», actitud común a Wittgenstein y Dewey. La mejor forma de entender esta clase de conductismo es considerarlo como una especie de holismo —pero un holismo que no necesita apuntalamientos metafísicos idealistas. Afirma que si entendemos las reglas de un juego lingüístico, entendemos todo lo que hay que entender sobre las causas por las que se hacen los movimientos en ese juego lingüístico (es decir, todo menos la comprensión extra obtenida de investigaciones que nadie llamaría epistemológicas —por ejemplo, de la historia del lenguaje, la evolución de la especie y el ambiente político o cultural de los jugadores). Si somos conductistas en este sentido, no se nos ocurrirá invocar ninguna de las tradicionales distinciones kantianas. Pero, ¿podemos seguir adelante y ser conductistas? O, como sugieren los críticos de Quine y Sellars, ¿no es el conductismo una simple petición de principio? [10]. ¿Existe alguna razón para pensar que las nociones epistemológicas fundamentales *deban* explicarse en términos conductistas?

---

[8] Véase una interpretación de Quine en cuanto ataque a la utilidad explicativa de la «idea filosófica de significado» en Gilbert Harman, «Quine on Meaning and Existence, I», *Review of Metaphysics*, 21 (1967), págs. 124-252, especialmente págs. 125, 135-141.

[9] Wilfrid Sellars, *Science, Perception and Reality* (Londres y Nueva York, 1963), pág. 167.

[10] Para esta clase de crítica del conductismo de Quine, véase H. P. Grice y P. F. Strawson, «In Defense of a Dogma», *Philosophical Review*, 65 (1956), páginas 141-156. Para críticas semejantes a Sellars, véase la crítica de Roderick Chisholm a sus afirmaciones sobre la intencionalidad, en su correspondencia publicada en *Minnesota Studies in the Philosophy of Science*, 2 (1958), páginas 521 y ss.

Esta última pregunta se reduce a la siguiente: ¿Podemos tratar el estudio de «la naturaleza del conocimiento humano» sólo como el estudio de ciertas formas en que se interactúan los seres humanos, o requiere un fundamento ontológico (que implica una forma específicamente filosófica de describir los seres humanos)? ¿Consideraremos que «S sabe que p» (o «S sabe no-inferencialmente que p», o «S cree incorregiblemente que p», o «El conocimiento de S de que p es cierto») es una observación sobre la categoría de los informes de S entre sus semejantes, o que es una observación sobre la relación entre sujeto y objeto, entre la naturaleza y su espejo? La primera alternativa nos lleva a una visión pragmática de la verdad y a un enfoque terapéutico de la ontología (en que la filosofía puede resolver las disputas absurdas entre sentido común y ciencia, pero no aportar ningún argumento propio en favor de la existencia o inexistencia de algo). Así, para Quine, una verdad necesaria es simplemente una afirmación tal que nadie nos ha dado ninguna alternativa interesante que nos lleve a ponerla en duda. Para Sellars, decir que un informe de un pensamiento pasajero es incorregible es decir que nadie ha sugerido todavía una forma de predecir y controlar la conducta humana que no acepte a pie juntillas los informes de pensamientos cuando son coincidentes en el tiempo, en primera persona y sinceros. La segunda alternativa lleva a explicaciones «ontológicas» de las relaciones entre mentes y significados, mentes y datos inmediatos de conciencia, universales y particulares, conciencia y cerebros, y así sucesivamente. Para filósofos como Chisholm y Bergmann, *hay que* intentar tales explicaciones, si queremos mantener el realismo del sentido común. El objetivo de todas estas explicaciones es hacer de la verdad algo más de lo que Dewey llamaba «afirmabilidad justificada»: más de lo que nuestros semejantes, *ceteris paribus,* nos dejarán que consigamos decir. Estas explicaciones, cuando son ontológcas, suelen tomar la forma de una redescripción del objeto de conocimiento para «llenar el vacío» entre él y el sujeto que conoce. Elegir entre estos enfoques es elegir entre la verdad como «lo que es bueno que creamos» y la verdad como «contacto con la realidad».

Así, la cuestión de si podemos ser conductistas en nuestra actitud hacia el conocimiento no es algo que dependa de la «adecuación» de los «análisis» conductistas de las pretensiones de conocimiento o de los estados mentales. El conductismo epistemológico (que podría llamare simplemente «pragmatismo», si no fuera éste un término algo sobrecargado) no tiene nada que ver con Watson o con Ryle. Más bien, es la afirmación de que la filosofía no tendrá nada

más que ofrecer que el sentido común (con la ayuda de la biología, la historia, etc.) en relación con el conocimiento y la verdad. La pregunta no es si se pueden ofrecer condiciones conductuales necesarias y suficientes para «S sabe que p»; nadie sueña con esa posibilidad. Tampoco se trata de saber si se pueden ofrecer tales condiciones para «S ve que p», o «A S le parece como si p», o «S está teniendo el pensamiento de que p». Ser conductista en el sentido amplio en que lo son Sellars y Quine no es ofrecer análisis reduccionistas, sino negarse a intentar cierto tipo de explicación: el tipo de explicación que no sólo interpone entre el impacto del entorno sobre los seres humanos y sus informes sobre él una noción como la «familiaridad con los significados» o «familiaridad con las apariencias sensoriales», sino que utiliza estas nociones para explicar la fiabilidad de tales informes.

Pero, una vez más, ¿cómo podemos determinar si son necesarias estas nociones? Resulta tentador contestar basándose en una decisión antecedente sobre la naturaleza de los seres humanos —una decisión sobre si necesitamos ideas como la de «mente», «curso de la conciencia» y otras semejantes para describirlas. Pero esto sería la respuesta fuerte. Podemos adoptar la actitud de Sellars-Quine hacia el conocimiento, y al mismo tiempo «aprobar» alegremente las sensaciones primarias, conceptos a priori, ideas innatas, datos sensoriales, proposiciones, y cualquiera otra cosa que una explicación causal de la conducta humana considerara útil postular [11]. Lo que *no podemos* hacer es considerar el conocimiento de estas realidades «internas» o «abstractas» como *premisas* de las que se infiere normalmente nuestro conocimiento de otras realidades, y sin las cuales este último conocimiento «carecería de base». La diferencia está entre decir que saber una lengua es estar familiarizado con los significados de sus términos, o que ver una mesa es tener una impresión sensorial rectangular, y explicar la *autoridad* de casos de «todos los hombres son animales» o «eso parece una mesa» en virtud de la autoridad previa (interna, privada, no social) de un conocimiento de los significados o de las impresiones sensoriales. En epistemología el conductismo no es cuestión de frugalidad metafísica, sino de si se puede

[11] Defiendo esta afirmación al hablar de la psicología empírica en el capítulo quinto. Los propios Sellars y Quine, por desgracia, no ven las cosas con tanta despreocupación. Véase una crítica a la huida de las intenciones por parte de Quine en la sección 3, más adelante. Esta crítica se ha aplicado, *mutatis mutandis*, a la insistencia de Sellars en la afirmación de que la «imagen científica» excluye las intenciones; pero lo que dice Sellars es más sutil y está relacionado con su idea tractaria de la imagen, criticada más adelante, en el capítulo sexto, sección 5.

atribuir autoridad a las afirmaciones en virtud de relaciones de «familiaridad» entre personas y, por ejemplo, pensamientos, impresiones, universales y proposiciones. La diferencia entre los puntos de vista de Quine-Sellars y de Chisholm-Bergmann en estas cuestiones no es la diferencia entre paisajes exuberantes y austeros, sino más bien algo parecido a la diferencia que hay entre los filósofos morales que piensan que los derechos y responsabilidades dependen de lo que decida la sociedad y los que piensan que hay algo dentro del hombre que la sociedad «reconoce» al tomar sus decisiones. Las dos escuelas de filosofía moral no se diferencian en el punto de que los seres humanos tengan derechos por los que vale la pena dar la vida. Se distinguen más bien en si, una vez que hemos comprendido cuándo y por qué se han concedido o negado estos derechos, de la forma en que entienden esto los historiadores sociales e intelectuales, queda algo más por entender. En pocas palabras, se distinguen al hablar de si hay «fundamentos ontológicos para los derechos humanos», igual que el enfoque de Sellars-Quine se distingue de las tradiciones empiristas y racionalistas al hablar de si, una vez que entendemos (como hacen los historiadores del conocimiento) cuándo y por qué se han adoptado o rechazado varias creencias, queda todavía por entender algo que se llama «la relación del conocimiento con la realidad».

Esta analogía con la filosofía moral nos permite centrar la atención una vez más en el problema del conductismo en la epistemología: el problema no es la adecuación de la explicación de hecho, sino más bien si se puede ofrecer de hecho una «base» a una práctica de justificación. La cuestión no es si el conocimiento humano tiene de hecho «fundamentos», sino si tiene sentido sugerir que los tiene —si la idea de que la autoridad epistémica o moral tenga una «base» en la naturaleza resulta una idea coherente. Para los pragmáticos en el campo de la moral, la afirmación de que las costumbres de una determinada sociedad están «basadas en la naturaleza humana» no es una afirmación de la que sepa argumentar. Es pragmático porque no ve qué puede ser eso de que una costumbre esté basada. Para el enfoque apistemológico de Quine-Sellars, decir que la verdad y el conocimiento sólo pueden ser juzgados por los criterios de los investigadores de nuestros días no es decir que el conocimiento humano sea menos noble o importante, o más «alejado del mundo», de lo que habíamos pensado. Es decir, simplemente, que nada figura como justificación a no ser por referencia a lo que aceptamos ya, y que no hay forma de salir de nuestras

creencias y de nuestro lenguaje para encontrar alguna prueba que no sea la coherencia.

Decir que la Verdad y el Bien son cuestiones de práctica social parece que nos condena a un relativismo que, por sí solo, es una *reductio* de un enfoque conductista o a conocimiento o a moralidad. Examinaré esta acusación al hablar del historicismo, en los capítulos siete y ocho. Aquí sólo quiero señalar que únicamente la imagen de una disciplina —la filosofía— que descubra que un conjunto determinado de opiniones científicas o morales es más «racional» que las que se presentan como alternativa, y hacerlo recurriendo a algo que forme una matriz neutra permanente para toda investigación y toda historia, hace posible pensar que dicho relativismo debe excluir automáticamente las teorías de coherencia de la justificación intelectual y práctica. Una razón por la que los filósofos profesionales rechazan la afirmación de que el conocimiento puede no tener fundamentos, o de que los derechos y deberes quizá no tengan una base ontológica, es que aquel conductismo que prescinda de los fundamentos lleva camino de prescindir de la filosofía. La opinión de que no hay una matriz neutra permanente dentro de la cual se representen los dramas de la investigación y la historia tiene el corolario de que la crítica de la propia cultura sólo puede ser fragmentaria y parcial —nunca «por referencia a criterios eternos». Es una amenaza para la imagen neo-kantiana de la relación de la filosofía con la ciencia y con la cultura. El deseo apremiante de decir que las afirmaciones y acciones sólo deben tener coherencia con otras afirmaciones y acciones, pero deben «corresponder» a algo aparte de lo que las personas están diciendo y haciendo tiene cierto derecho a ser llamado *el* impulso filosófico. Es el impulso que llevó a Platón a decir que las palabras y hechos de Sócrates, a pesar de que eran coherentes con la teoría y práctica del tiempo, correspondían, sin embargo, a algo que los atenienses no podían ni siquiera imaginar. El platonismo residual a que se oponen Quine y Sellars no es la hipostatización de realidades no físicas, sino la idea de la «correspondencia» con tales realidades como piedra de toque que sirva para determinar el valor de la práctica actual [12].

---

[12] Por desgracia, los dos suelen utilizar la correspondencia en vez de las entidades físicas, y específicamente en vez de las «entidades básicas» de la ciencia física (partículas elementales, o sus sucesores). El intento de Sellars (y de Jay Rosenberg) de salvar *algo* de la idea platónica del conocimiento en cuanto exactitud de la imagen será criticado más adelante (capítulo sexto, sección 5). Mi propia actitud es la de Strawson (y Heidegger): «La teoría de la correspondencia necesita no ser purificada sino eliminada» (P. F. Strawson, «Truth», reimpreso en *Truth,* edición George Pitcher [Englewood Cliffs,

En resumen, lo que estoy afirmando es que el ataque de Quine-Sellars a la idea kantiana de dos clases de representaciones —intuiciones «dadas» a una facultad, y conceptos (o significados) «dados» a otra— no es el intento de sustituir una forma de explicación del conocimiento humano por otra, sino el intento de abandonar la noción de «una explicación del conocimiento humano». Equivale a una protesta contra un problema filosófico arquetípico: el problema de cómo reducir las normas, reglas y justificaciones a hechos, generalizaciones y explicaciones [13]. Por esta razón, no encontraremos una base metafilosófica neutra en que apoyarnos para demostrar las cuestiones que plantean Quine y Sellars. No están ofreciendo una «explicación» cuya «adecuación» haya que comprobar, sino señalando la inutilidad de presentar una «explicación». Negarse, como hacen ambos, a justificar las afirmaciones recurriendo a episodios conductistamente inverificables (en que la mente reconoce su propia familiaridad directa con una instanciación de verdor o con el significado de «verde») es sencillamente decir que la justificación ha de ser holística. Si queremos tener una doctrina del «conocimiento por familiaridad» que nos ofrezca un fundamento, y si no nos limitamos a negar que haya justificación, entonces afirmaremos, con Sellars, que «la ciencia es racional no porque tenga un *fundamento,* sino porque es una empresa que se auto-corrige y puede hacer peligrar *cualquier* afirmación, aunque no *todas* a la vez» [14]. Diremos, con Quine, que el conocimiento no es como una estructura arquitectónica sino como un campo de fuerza [15], y que no hay afirmaciones que sean inmunes a la revisión. Seremos holistas no porque tengamos afición a los conjuntos, lo mismo que no somos conductistas porque nos desagraden las «realidades fantasmagóricas», sino sim-

---

N. J., 1964], pág. 32) —o, dicho más suavemente, necesita ser separada de la epistemología y relegada a semántica. (Véase Robert Brandom, «Truth and Assertability», *Journal of Philosophy,* 73 [1976], págs. 137-149.)

[13] Cfr. la afirmación de Sellars de que «la idea de que los hechos epistémicos pueden ser analizados en su totalidad —incluso 'en principio'— en hechos no epistémicos, sean fenoménicos o conductuales, públicos o privados, quizá con una generosa profusión de subjuntivos e hipotéticas, es, creo yo, un error radical —un error semejante a la llamada 'falacia naturalista' en ética» *(Science, Perception and Reality,* pág. 131). Yo diría que la importancia del enfoque epistemológico de Sellars es que ve la irreductibilidad verdadera e interesante en esta área no como si fuera entre una clase de particulares (mentales, intencionales) y otra (física), sino entre descripciones, por un lado, y normas, prácticas y valores, por otro. (Véase la nota 17 más adelante.)

[14] Sellars, *Science, Perception and Reality,* pág. 170.

[15] W. V. O. Quine, *From a Logical Point of View* (Cambridge, Mass., 1953), página 42.

plemente porque la justificación ha *sido* siempre conductista y holística. Sólo el filósofo profesional ha soñado que podría haber alguna otra cosa, pues sólo él tiene miedo al escéptico epistemológico. Un planteamiento holístico del conocimiento no es una polémica anti-fundamentalista, sino fruto de la desconfianza en toda la empresa epistemológica. Un planteamiento conductista de los episodios de «conciencia directa» no es una polémica antimentalista, sino fruto de una desconfianza en la búsqueda platónica de esa clase especial de certeza asociada con la percepción visual. La imagen del Espejo de la Naturaleza —un espejo que se ve con más facilidad y certeza que lo que refleja— sugiere, y está sugerida por, la imagen de la filosofía en cuanto tal búsqueda.

Si lo que vengo diciendo hasta ahora es correcto, no hay forma posible de defender los puntos de vista de Sellars y Quine como no sea replicando a sus críticos. No hay ninguna base neutral en que apoyarse y demostrar que han vencido, respectivamente, a «lo dado» y a «lo analítico» en buena lid. Lo mejor que podemos hacer es separar la forma pura de sus críticas de la tradición y las diversas cuestiones ajenas que han introducido sus críticos (y, en cierta medida, los propios Quine y Sellars), y quizá reducir de esta manera el aspecto paradójico de sus doctrinas. En la siguiente sección, consideraré el ataque de Sellars al Mito de lo Dado, y trataré de separarlo de las implicaciones «desagradables para los niños» de la afirmación de que no hay eso que se llama conciencia pre-lingüística. A continuación, abordaré el ataque de Quine a la distinción entre lenguaje y hecho y trataré de separarlo de las desafortunadas afirmaciones reduccionistas de Quine al hablar de la «indeterminación» de la traducción y de las *Geisteswissenschaften*. Las doctrinas de Sellars y Quine, una vez purificadas, parecen expresiones complementarias de una sola afirmación: que ninguna «explicación de la naturaleza del conocimiento» puede basarse en una teoría de las representaciones que están en relaciones privilegiadas con la realidad. La obra de estos dos filósofos nos permite desenmarañar, por fin, la confusión de Locke entre explicación y justificación, y dejar claro por qué una «explicación de la naturaleza del conocimiento» puede ser, en el mejor de los casos, una descripción de la conducta humana.

3. Conciencia pre-lingüística

En «Empiricism and the Philosophy of Mind», Sellars formula el «nominalismo psicológico» como la concepción de que

> *toda* conciencia de *clases, parecidos, hechos,* etc., en resumen toda conciencia de entidades abstractas —en realidad, toda conciencia, hasta de particulares— es una cuestión lingüística. De acuerdo con esto, ni siquiera la conciencia de las clases, parecidos y hechos que pertenecen a la llamada experiencia inmediata está presupuesta por el proceso de adquirir el uso del lenguaje [16].

La existencia de sensaciones primarias —dolores, las sensaciones que puedan tener los bebés cuando miran objetos de colores, etcétera— es la objeción obvia a esta doctrina. Para responder a esta objeción, Sellars recurre a la distinción entre la conciencia-como-conducta-discriminatoria y la conciencia como lo que Sellars llama estar «en el espacio lógico de las razones, de justificar y ser capaz de justificar lo que se dice» (pág. 169). La conciencia en el primer sentido se manifiesta en las ratas y amebas y computadores; es sencillamente hacer señales de forma fiable. La conciencia en el segundo sentido se manifiesta únicamente en los seres cuya conducta interpretamos como la emisión de oraciones con la intención de justificar la emisión de otras oraciones. En este último sentido la conciencia es la creencia verdadera justificada —conocimiento— pero en el primer sentido es la capacidad de responder a los estímulos. Gran parte de la «Epistemology and the Philosophy of Mind» es una demostración de que esta capacidad es condición causal para el conocimiento pero no una *base* para el conocimiento. Esta concepción tiene el corolario de que el conocimiento de los particulares o de los conceptos no es anterior, temporalmente, al conocimiento de las proposiciones (sino siempre una abstracción de éstas), y, por tanto, que las explicaciones empiristas del aprendizaje del lenguaje y de la base no proposicional del conocimiento están descaminadas inevitablemente. La premisa fundamental de este argumento es que no existe ninguna creencia justificada que sea no-proposicional, y ninguna justificación que no sea una relación entre proposiciones. Siempre será un error hablar de nuestro contacto con el verdor o con una instanciación del verdor como si fueran la «base» (en oposición a ser una condición causal) de nuestro conocimiento de que «esto es un objeto verde» o que el «verdor es un color».

Tanto los niños como las células fotoeléctricas discriminan los objetos verdes, pero se piensa que los niños pre-lingüísticos «saben qué es verde» en un sentido en que las células fotoeléctricas no lo saben. Pero, ¿cómo puede saber el niño lo que es un dolor si

[16] Sellars, *Science, Perception and Reality,* pág. 160.

todo conocimiento de cualquier cosa «es una cuestión lingüística»? En esto Sellars necesita otra distinción. Esta vez se trata de la distinción entre «saber cómo es X» y «saber qué clase de cosa es X». Esto último implica ser capaz de vincular el concepto de Xeidad con otros conceptos de forma que se puedan justificar las afirmaciones sobre Xs. Según la concepción wittgensteiniana de Sellars, en la cual tener un concepto es usar una palabra, estas dos capacidades son una misma capacidad. Se deduce que no podemos tener un concepto sin tener muchos, ni podemos llegar «a tener un concepto de algo porque hayamos observado ese tipo de cosa»; pues «tener la capacidad de observar un tipo de cosa es ya tener el concepto de este tipo de cosa» (pág. 176). Pero «observar un tipo de cosa» es observar bajo una descripción, no sólo responder discriminativamente a ella. ¿Qué es, entonces, saber *cómo* es el dolor sin saber u observar qué *tipo* de cosa es?

Es sencillamente *tener* dolor. La trampa que debemos evitar en este punto es la idea de que hay alguna iluminación interna que sólo tiene lugar cuando la mente del niño está iluminada por el lenguaje, conceptos, descripciones y proposiciones, y no tiene lugar cuando el niño sufre y llora inarticuladamente. El niño *siente* la misma cosa, y siente *lo mismo* antes y después del aprendizaje del lenguaje. Antes del lenguaje se dice que *conoce* la cosa que siente por si acaso es la misma clase de cosa de la que, más adelante, será capaz de hacer informes no inferenciales. Es esta capacidad latente lo que le separa de la célula fotoeléctrica, no su mayor sensibilidad. Así, puede responder directamente a una ausencia de oxígeno en el aire, al movimiento demasiado rápido de las moléculas, a la alteración de los ritmos alfa de su cerebro, y así sucesivamente, pero no se dice que «sepa lo que son» a no ser que y mientras no haya conseguido dominar el vocabulario relevante. Pero el calor, sofoco, éxtasis, dolor, fuego, verdor, hostilidad paterna, amor materno, hambre, intensidad del sonido, y cosas semejantes, son «conocidos» prelingüísticamente, o al menos eso diría el habla ordinaria. Se conocen porque se han *tenido* o sentido. Son conocidos sin que puedan ser colocados en clases, o relacionados de ninguna otra forma con ninguna otra cosa.

No hay razón para que Sellars ponga objeciones a la idea de «saber cómo es el dolor (o el verdor)», pues esto sólo serviría para confirmar el Mito de lo Dado, y estaría en contra del nominalismo psicológico, si hubiera conexión entre saber cómo es el dolor y saber qué clase de cosa es el dolor. Pero la única conexión es que lo primero es una condición causal insuficiente e innecesaria para

lo segundo. Es insuficiente por la razón evidente de que podemos saber cómo es el verdor sin saber que es diferente del azul, que es un color, etc. Es innecesario porque podemos saber todo eso, y muchas cosas más, sobre el verdor aun siendo ciegos de nacimiento, y, por tanto, si *no* supiéramos cómo es el verdor. Es falso que no podamos hablar ni saber de lo que no tenemos sensaciones primarias, e igualmente falso que si no podemos hablar de ellas podemos, sin embargo, tener creencias verdaderas justificadas sobre ellas. Lo que hay de especial en el lenguaje no es que «cambie la cualidad de nuestra experiencia» o «abra nuestras perspectivas a la conciencia» o «sintetice una diversidad previamente inconsciente» o produzca cualquier otro tipo de cambio «interior». Lo único que hace su adquisición es dejarnos entrar en una comunidad cuyos miembros intercambian justificaciones y afirmaciones, y otras acciones, entre sí [17].

Podemos considerar que Sellars se dirige al empirismo tradicional con estas palabras: el saber cómo son las cosas no es cuestión de estar justificado al afirmar proposiciones. A esto, el empirista respondería probablemente, como han hecho Roderick Firth y otros, que esta concepción confunde los conceptos con las palabras [18]. Se

---

[17] Cfr. *ibíd.*, pág. 169: «Lo esencial es que al caracterizar un episodio o estado como de *conocer,* no estamos dando una descripción empírica de dicho episodio o estado; lo estamos colocando en el espacio lógico de las razones, de justificar y ser capaz de justificar lo que se dice.»

[18] Véase Roderick Firth, «Coherence, Certainty, and Epistemic Priority», reimpreso en *Empirical Knowledge,* ed. Roderick Chisholm y Robert Swartz (Englewood Cliffs, N. J., 1973). Firth trata de resolver un problema creado por la «teoría de la coherencia de los conceptos»: a saber, dado que «no podemos entender totalmente 'parece rojo' si no poseemos el concepto de contraste 'es rojo', parecería que *no* es lógicamente posible tener el concepto de 'parece rojo' [como requería la doctrina de C. K. Lewis del 'significado sensorial'] antes de tener el concepto 'es rojo'» (pág. 461). Firth dice que

> la paradoja subyacente se elimina fácilmente si no confundimos los conceptos con las palabras utilizadas para expresarlos. Es un hecho genético, pero un hecho que tiene implicaciones filosóficas, que cuando un niño comienza a utilizar por primera vez la palabra «rojo» con cierta coherencia la aplica a cosas que le *parecen* rojas... Llamar a esto una «forma primitiva» del concepto «parece rojo» es reconocer que en cierto sentido el niño no puede entender *totalmente* el uso adulto hasta que no sea capaz de distinguir las cosas que solamente parecen rojas de las que realmente lo son; pero no debemos suponer que el niño *pierda* de alguna manera su concepto primitivo cuando adquiere otro más elaborado (págs. 461-462).

Véase una discusión más detallada de la identificación de Ryle-Wittgenstein-Sellars entre tener conceptos y utilizar palabras, en Brand Blanshard, *Reason*

dice que Sellars, Wittgenstein y otros que «exageran» la importancia del lenguaje incurren en una petición de principio en favor del nominalismo psicológico al suponer que tener un concepto es tener el uso de una palabra. Sellars puede replicar con el siguiente dilema: o atribuís conceptos a todo aquello (por ejemplo, los cambiadiscos) que puede responder discriminadamente a clases de objetos, o explicad por qué trazáis la línea divisoria entre pensamiento conceptual y sus predecesores primitivos en un lugar diferente del que separa haber adquirido un lenguaje y estar todavía en proceso de adiestramiento. Este dilema llama la atención sobre el hecho de que las ideas tradicionales de lo dado han unido las sensaciones primarias y la capacidad de discriminar, utilizando la ausencia de las primeras para eliminar a las máquinas e incluir a los bebés, y utilizando luego la presencia de la segunda para hacer que lo que tienen los niños se parezca al conocimiento proposicional. La discusión entre Sellars y sus críticos se reduce en este punto a lo siguiente: ¿Vamos a considerar la conceptualización como una cuestión de clasificación o de justificación? Sellars puede decir que cederá el término *concepto* a quienes deseen dotar a los cambiadiscos o a sus parejas protoplásmicas de conceptos, con tal que él pueda tener algún otro término para indicar lo que tenemos cuando podemos colocar clasificaciones en relación con otras clasificaciones de la forma que hacen los usuarios de un lenguaje cuando discuten sobre cuál es la clase en que se debe incluir un ítem determinado. Una vez más, Sellars recurre a decir que la justificación es cuestión de práctica social, y que todo lo que no sea cuestión de práctica social no ayuda en absoluto a entender la *justificación* del conocimiento humano, independientemente de los servicios que pueda prestar para entender su *adquisición*. La falacia naturalista y la falacia genética, piensa Sellars, se han combinado en el empirismo tradicional para producir la idea de que estaríamos en mejores condiciones de felicitarnos por reflejar exactamente la naturaleza (o de lamentarnos por no conseguirlo) si pudiéramos al menos traer a la conciencia las etapas del desarrollo de nuestra infancia. Confundidos por la mezcla cartesiana de pensamiento y sensación, absortos ante la inocencia virginal de la tablilla de cera de Locke, y asustados por el hecho de que si la verdad está en el todo la certeza no está en ninguna parte, los empiristas se han aferrado a «cómo parece lo rojo» como

and Analysis (La Salle, Ill., 1962), capítulo IX. Algo parecido a la idea de Firth de un «concepto primitivo» reaparece en una crítica a Sellars por William S. Robinson, «The Legend of the Given», en *Action, Knowledge and Reality*, ed. H.-N. Castañeda (Indianapolis, 1975), págs. 83-108.

si fuera la clase de nuestro conocimiento del mundo natural. Para Sellars, esto es igual que aferrarse a cómo se encuentra el niño cuando tardan en darle de comer y considerarlo como la clave de la conciencia moral común.

Resumiendo: el nominalismo psicológico de Sellars no es una teoría de cómo funciona la mente, ni de cómo nace el conocimiento en el corazón del niño, ni de la «naturaleza de los conceptos», ni de ninguna otra cuestión de hecho. Es una observación sobre la diferencia entre hecho y reglas, con el propósito de que sólo podemos caer bajo reglas epistémicas cuando hayamos entrado en la comunidad donde se practica el juego gobernado por estas reglas. Podemos negarnos a afirmar que el conocimiento, la conciencia, los conceptos, el lenguaje, la inferencia, la justificación y el espacio lógico de las razones caigan sobre los hombros del niño despierto en algún momento situado en torno a los cuatro años de edad, sin que hasta ese momento hayan existido ni siquiera en la forma más primitiva posible. Pero no rechazamos el pensamiento de que al cumplir dieciocho años recaerán sobre él todo un conjunto de derechos y responsabilidades, sin que hasta entonces hayan estado presentes ni en la más primitiva de las formas. Esta última situación es, ciertamente, más clara y definida que la primera, pues no hay ninguna señal de la primera circunstancia como no sea en algún comentario incidental de un adulto (por ejemplo, «el niño sabe de qué está hablando»). Pero en ambos casos lo que ha ocurrido ha sido un cambio en las relaciones de una persona con las otras, no un cambio dentro de la persona que le *acomode* ahora a entrar en estas nuevas relaciones. No se trata de que podamos estar *equivocados* al pensar que un niño de cuatro años tiene conocimiento y no lo tenga ningún niño de un año, en mayor medida en que podamos estar *equivocados* al aceptar a pie juntillas lo que nos dice la legislación del hecho de que las personas de dieciocho años se pueden casar con toda libertad y los de diecisiete no pueden hacerlo. Puede ser *imprudente* tomarse en serio los balbuceos de un niño de cuatro años, lo mismo que puede haber sido poco sensato fijar la responsabilidad legal a una edad tan temprana, pero la mayor comprensión de cómo «funciona» el conocimiento (o la responsabilidad) no resolverá nunca estas cuestiones.

Por eso, no hay que esperar de Sellars que presente argumentos en favor de una «teoría de la relación entre lenguaje y pensamiento», pues los pensamientos son episodios internos que pueden entenderse o no entenderse (según las necesidades de la psicología empírica) como vinculados necesariamente con el lenguaje, los es-

tados cerebrales, o varias otras cosas. En cuanto *epistemólogo,* Sellars no está ofreciendo una teoría sobre los episodios internos. Más bien, está señalando que la idea tradicional, no conductista, de «epistemología» es la confusión de una explicación de dichos episodios con una explicación del derecho a hacer ciertas afirmaciones. Esto es adoptar el parecer de que la filosofía (y, específicamente, la «filosofía de la mente») no puede, por el hecho de presentar un punto de vista más elevado, reforzar o debilitar la confianza en nuestras propias afirmaciones que nos da la aprobación de nuestros semejantes. El nominalismo psicológico de Sellars no procede del conductismo en cuanto tesis sobre lo que es o deja de ser la mente. Procede únicamente del conductismo epistemológico en el sentido definido más arriba, sentido que resulta indistinguible del holismo epistemológico. Ser conductista en este sentido es simplemente dividir por cualquiera y por todos los hechos y facultades mentales y considerar que nuestras prácticas de justificar las afirmaciones no necesitan base empírica u «ontológica» [19].

Tras volver de nuevo a la comunidad en cuanto fuente de autoridad epistémica, terminaré esta sección insistiendo una vez más en que incluso el conocimiento no conceptual, no lingüístico, de cómo es una sensación se atribuye a los seres basándose en su posible inclusión dentro de esta comunidad. A los bebés y a los animales más atractivos se les atribuye el «tener sensaciones» y no (como las células fotoeléctricas y los animales por los que no sentimos inclinación alguna —por ejemplo, las platijas y arañas—) «responder meramente a los estímulos». Esto ha de explicarse basándose en esa especie de sentimiento comunitario que nos une con lo que es humanoide. Ser humanoide es tener una cara humana, y la parte más importante de esa cara es una boca que podemos imaginar profiriendo oraciones en sincronía con las adecuadas expresiones de la cara en su conjunto [20]. Decir, con el sen-

---

[19] En el capítulo octavo mantendré que esta actitud hacia la relación entre filosofía de la mente y epistemología (o, en sentido más amplio, entre cualquier descripción científica o metafísica del hombre, por una parte, y cualquier observación sobre la justificación de sus acciones —lingüísticas o de otra naturaleza—, por la otra) ocupa un lugar central en los esfuerzos que Wittgenstein realiza en las *Investigations.* Creo que, como la aplicación que hace Sellars de la distinción entre describir hombres y juzgarlos, es un corolario natural de la rigurosa separación del *Tractatus* entre la afirmación de hechos y todos los demás usos del lenguaje (por ejemlpo, éticos).

[20] Véase Virgil Aldrich, «On What It Is Like to Be a Man», *Inquiry,* 16 (1983), págs. 355-366; contiene una buena explicación de la fuerza y ramificaciones de la opinión de Wittgenstein de que el cuerpo humano es la mejor

tido común, que los bebés y los murciélagos saben cómo es el dolor y el verde, pero no cómo es el movimiento de las moléculas o el cambio de estaciones, es decir simplemente que podemos imaginárnoslos con relativa facilidad abriendo la boca y haciendo comentarios sobre lo primero, pero no sobre lo segundo. Decir que un aparato (formado por una célula fotoeléctrica conectada a una cinta magnetofónica) que diga «¡verde!» cuando, y sólo cuando, proyectemos sobre él una luz verde *no* sabe cómo es el color verde es decir que no podemos imaginarnos fácilmente cómo mantener una conversación con el aparato. Decir que no *sabemos* si los androides producidos a partir de protoplasma (y totalmente preparados para funcionar, si no fuera por el centro de velocidad que está a punto de ser instalado) saben cómo es el verde es no confesar el desconcierto científico o filosófico en torno a la naturaleza de la subjetividad[21]. Es decir, simplemente, que a las cosas que tienen rostros más o menos humanos y parece como si algún día pudieran ser interlocutores conversacionales se les atribuyen «sensaciones», pero que si sabemos con detalle cómo se han montado estos aparatos podemos ser reacios a pensar en ellos como posibles interlocutores[22].

Esta concepción de la atribución de conciencia pre-lingüística —como cumplido tributado a los posibles o imaginarios colegas que comparten nuestro lenguaje— tiene un corolario: las prohibiciones morales de hacer daño a los niños y a los animales de aspecto más agradable no están «basadas ontológicamente» en su posesión de sensaciones. En todo caso, sería al revés. Las prohibiciones morales son expresiones de un sentido de comunidad basado en la posibilidad imaginaria de conversación, y la atribución de sensaciones es poco más que un recuerdo de estas prohibiciones. Podemos ver esto si nos fijamos en que a nadie, como no sean los filósofos de

imagen del alma humana. Véase también Stuart Hampshire, *Thought and Action* (Londres, 1959), capítulo 1.

[21] Véase Hilary Putnam, «Robots: Machines or Artificially Created Life?», reimpreso en *Modern Materialism,* ed. J. O'Connor (Nueva York, 1969), página 262: «... la pregunta: ¿Son conscientes los robots? exige una decisión, por nuestra parte, de tratar a los robots como colegas de nuestra comunidad lingüística, o no tratarlos así.»

[22] Esto no significa que tengamos razón para aborrecerlo, ni que no la tengamos. Simplemente quiero llamar la atención sobre el miedo tradicional a que los biólogos y psicólogos, en su actitud hacia sus semejantes, puedan «matar para hacer disecciones». Tenemos la tentación, cuando somos especialmente dotados para prever la conducta de algo basándonos en su estructura interna, de ser «obejtivos» al respecto —es decir, de tratarlo como un *en-soi* en vez de como un *pour-soi* y «uno de nosotros». Creo que los filósofos no

la mente, le importa si la sensación primaria de dolor o verdor es diferente en el caso de los koalas o de nosotros, pero todos nos interesamos algo por un koala cuando le vemos retorcerse de dolor. Este hecho no significa que nuestro dolor o el del koala «no sea más que su conducta»; significa simplemente que para nuestra capacidad de imaginar que el koala nos está pidiendo ayuda su contorsión de dolor es más importante que lo que ocurre dentro del koala. Los cerdos ocupan un puesto mucho más elevado que los koalas en los tests de inteligencia, pero los cerdos no se contorsionan de forma muy humanoide, y su cara no tiene la forma adecuada para las expresiones faciales que acompañan a la conversación normal. Por eso vemos sacrificar a los cerdos sin inmutarnos, y en cambio formamos sociedades para la protección de los koalas. Esto no es «irracional», lo mismo que no es irracional ampliar o negar los derechos civiles a los subnormales (o a los fetos, o a las tribus aborígenes, o a los marcianos). La racionalidad, considerada como la formación de silogismos basados en el descubrimiento de «los hechos» y en la aplicación de principios como «El dolor debe reducirse al mínimo» o «La vida inteligente es siempre más valiosa que los seres bellos no inteligentes», es un mito. Sólo el impulso platónico de decir que todo sentimiento moral, y en realidad toda emoción de cualquier clase, debe estar basado en el reconocimiento de una cualidad objetiva del receptor nos hace pensar que nuestro tratamiento de los koalas o de los blancos o de los marcianos es una «cuestión de principio moral». Los «hechos» que se deben descubrir para aplicar el principio son, en el caso de las «sensaciones» del koala o del blanco, algo que no se puede descubrir independientemente del sentimiento [23]. Las emociones que tenemos hacia los casos fronterizos dependen de la viveza de nuestra imaginación, y al revés. Unicamente la idea de que en la filo-

---

tienen mucho que decir sobre cuándo esta tentación es justificada o injustificada. En cambio, los novelistas y poetas, sí.

[23] Es notorio que los filósofos morales no son de mucha ayuda para determinar qué es lo que debe considerarse como agente moral, como dotado de dignidad y no de valor, como uno de los seres cuya felicidad hay que conseguir al máximo, etc. Sellars considera brevemente el tema y se pregunta si todos los miembros de una comunidad epistémica son miembros de una comunidad ética —si «la intención intersubjetiva de promover el bienestar epistémico implica la intención intersubjetiva de promover el bienestar *sans phrase*» (*Science and Metaphysics* [Londres y Nueva York, 1968], pág. 225). Sobre la influencia del holismo en la meta-ética y en el impulso platónico que acabamos de mencionar, véase J. B. Schneewind, «Moral Knowledge and Moral Principles», en *Knowledge and Necessity,* ed. G. A. Vesey (Londres y Nueva York, 1970).

sofía tenemos una disciplina capaz de dar razones de peso de lo que creemos por instinto nos permite pensar que un «análisis filosófico más atento» nos ayudará a trazar una línea de separación entre la frialdad de corazón y el sentimentalismo absurdo.

Esta afirmación de que el conocimiento por parte de los animales de cómo son algunas cosas tiene poco que ver con una creencia verdadera justificada, y mucho que ver con la moral, se deduce naturalmente de la idea sellarsiana de que el interior de las personas y cuasi-personas debe explicarse por lo que ocurre en el exterior (y, en especial, por su lugar en nuestra comunidad) y no al revés. Desde que Descartes hiciera del solipsismo metodológico la señal distintiva del pensamiento filosófico riguroso y profesional, los filósofos han querido encontrar la «base» del conocimiento, de la moralidad, del gusto estético, y de todo lo que tiene importancia dentro del individuo. ¿Cómo es posible que haya en las sociedades algo que los individuos no hayan puesto allí? Sólo desde Hegel han comenzado los filósofos a jugar con la idea de que el individuo separado de su sociedad no es más que un animal. Las implicaciones antidemocráticas de esta concepción, por no decir nada de sus implicaciones historicistas y relativistas, han impedido que los modos de pensamiento hegelianos hayan producido impacto en la misma entraña de la filosofía analítica —la epistemología, la filosofía del lenguaje y la filosofía de la mente. Pero la obra «Empiricism and the Philosophy of Mind», de Sellars, descrita por el mismo autor como *«Meditations Hegeliènnes* incipientes» [24], consigue separar las sensaciones primarias y la creencia verdadera justificada, y privar a las sensaciones primarias de su condición de representaciones privilegiadas. Con ello demuestra cómo el conductismo epistemológico puede evitar la confusión entre explicación y justificación que hizo que la epistemología empirista pareciera posible y necesaria. En los capítulos siete y ocho trataré de demostrar cómo la insistencia en la prioridad de lo público ante lo privado, que es consecuencia de rechazar el empirismo, prepara el terreno a nuevos proyectos hegelianos y heideggerianos de destrucción.

4. La «idea 'idea'»

Tras demostrar que el ataque de Sellars al Mito de lo Dado es compatible con la bondad hacia los niños y animales y, por

[24] Sellars, *Science, Perception and Reality,* pág. 148.

tanto, con la conciencia moral común, quiero pasar ahora a demostrar que el ataque de Quine a la «idea 'idea'» y la distinción entre lenguaje y hecho es compatible con la respetabilidad intelectual de las *Geisteswissenschaften*. Las doctrinas de Quine sobre la «indeterminación de la traducción» y la «inescrutabilidad de la referencia» le han llevado a afirmar que no hay ninguna «cuestión de hecho» implicada en las atribuciones de significado a las elocuciones, de creencias a las personas, y de aspiraciones a las culturas. Creo que también en este caso algunas distinciones se encargarán de eliminar las supuestas consecuencias contra-intuitivas del conductismo epistemológico, haciendo posible que lo veamos como una preparación del terreno para la moralidad y la cultura superior y no como si las privara de la «verdad objetiva».

Lo que Quine llama la «idea 'idea'» es la concepción de que el lenguaje es expresión de algo «interior» que debe descubrirse antes de que podamos saber qué significa una elocución, o interpretar la conducta lingüística de los que realizan las elocuciones (por ejemplo, atribuirles creencias, deseos y culturas). Abandonar esta idea es abandonar simultáneamente la idea empirista-lógica de la «verdad en virtud del significado» y la idea oxoniana de la «verdad conceptual», pues no hay significados o conceptos donde puedan leerse. Esta actitud hacia el concepto de «concepto» hace posible prescindir de la distinción de Kant entre verdades necesarias (que pueden determinarse mirando sólo los conceptos [verdad analítica] o los conceptos puros y formas puras de la intuición únicamente [verdad sintética a priori]) y verdades contingentes (que requieren una referencia a las intuiciones empíricas). Pero Quine considera los conceptos y significados solamente como una especie de intenciones, y su deseo es borrar *todas* las intenciones. Así, admitiendo que «significa», «cree» y «desea», por ejemplo, no tienen equivalentes conductistas (como Brentano y Chisholm, en un esfuerzo por conservar un núcleo de verdad en el dualismo tradicional mente-cuerpo, habían intentado demostrar), Quine concluye que esto demuestra que las nociones de «creencia» y «deseo» son (a efectos «científicos») tan innecesarias como las de «concepto» e «intuición»:

> Se puede interpretar que la tesis de Brentano demuestra o bien el carácter indispensable de los giros intencionales y la importancia de una ciencia autónoma de la intención, o bien lo infundado de los giros intencionales y la carencia de contenido de una ciencia de la intención. Mi postura, a diferencia de Brentano, es la segunda. Aceptar a pie juntillas el uso inten-

> cional es, como vimos, postular que las relaciones de traducción son algo objetivamente válido aunque indeterminado en principio en relación a la totalidad de las disposiciones de habla. Esta postulación no promete muchas ventajas en cuanto a penetración científica si no encontramos mejor base que decir que las supuestas relaciones de traducción están presupuestas por la lengua vernácula de la semántica y la intención[25].

Quine cree que este anti-intencionalismo es semejante a su polémica contra la analicidad. Pero no lo es. El autor de «Two Dogmas of Empiricism» *debería* haber dicho que los conceptos y significados son inofensivos si se postulan para ofrecer explicaciones de nuestra conducta, y se vuelven nocivos únicamente cuando se tratan como fuente de una clase especial de verdad y de una clase especial de autoridad para ciertas afirmaciones. En concreto, esperaríamos que dijera que las razones presentadas normalmente para traducir idiomas en un sentido y no en otro (o para adscribir un conjunto de creencias y deseos en vez de una alternativa cualquiera que predijera la misma conducta lingüística) están justificadas simplemente por su coherencia interna, y que las prácticas como la traducción y la adscripción de estados intencionales están justificadas por su utilidad social. Quine reconoce la utilidad, pero considera filosóficamente importante insistir en que la clase de verdad proferida en comentarios como «'Hund' es el equivalente alemán de 'perro'», y «Robinson cree en Dios» no es la clase de verdad que expresa «una cuestión de hecho»[26]. Nos brinda, así, una distinción entre verdad por conveniencia y verdad por correspondencia, por así decirlo, en vez de la antigua distinción positivista entre verdad por convención y verdad confirmada por experiencia sensorial. Las verdades sobre significados y creencias y proposiciones no son realmente verdades en el pleno sentido del tér-

[25] W. V. O. Quine, *Word and Object* (Cambridge, Mass., 1960), pág. 221.

[26] Véase *Words and Objections: Essays on the Work of W. V. Quine*, ed. Donald Davidson y Jaakko Hintikka (Drodrecht, 1969), pág. 303, donde Quine dice:

> Pensemos... en la totalidad de las verdades de la naturaleza, conocidas y desconocidas, observables e inobservables, pasadas y futuras. Lo que interesa de la indeterminación de la traducción es que resiste incluso a toda esta verdad, a la verdad total de la naturaleza. Esto es lo que quiero decir cuando digo que, donde se aplica la indeterminación de la traducción, no se da realmente el caso de una opción correcta; no ocurre nada semejante ni siquiera *dentro de* la infradeterminación reconocida de una teoría de la naturaleza.

mino —igual que los positivistas decían que las verdades necesarias no hablaban realmente «sobre el mundo».

El holismo y el pragmatismo de «Two Dogmas» parecería hacer esta distinción entre dos clases de verdad tan difícil de mantener como las antiguas distinciones que Quine se propone atacar. Muchos críticos de Quine han señalado este punto, y han afirmado que su insistencia en esta distinción es un residuo del empirismo tradicional[27]. Estoy de acuerdo con la mayoría de estas críticas, y no voy a tratar de resumirlas o sintetizarlas (sólo señalaré que estas críticas coinciden en afirmar que cualquier tipo de «indeterminación» que podamos encontrar en la traducción aparecerá, con el mismo carácter inofensivo, en las *Naturwissenschaften).* Pero examinando la idea de «inescrutabilidad de la referencia» podemos conseguir cierta comprensión de las intuiciones «empiristas» que hacen que Quine siga hablando de «correspondencia», y no le dejan incidir en las implicaciones de su propio conductismo y holismo.

Quine resume el argumento de su «Ontological Relativity» diciendo:

> Lo que tiene sentido es decir no qué son los objetos de una teoría, hablando absolutamente, sino cómo una teoría de los objetos es interpretable o reinterpretable en otra... Estas reflexiones nos llevan a apreciar que el enigma de ver las cosas al revés, o en colores complementarios, merece tomarse en serio y aplicar sus conclusiones con profusión. La tesis relativista a que hemos llegado es ésta: no tiene sentido decir qué son los objetos de una teoría, como no sea para interpretar o reinterpretar esa teoría en otra... El hablar de las teorías subordi-

[27] La mejor conocida de estas críticas es la de Noam Chomsky, «Quine's Empirical Assumptions» en *Words and Objections.* Hilary Putnam insiste muy convincentemente en el tema en «The Refutation of Convencionalism», *Nous,* 8 (1974), 38: «Si la adopción de un sistema de hipótesis analíticas en vez de otro permite una mayor simplificación de ciencias como la neurofisiología, la psicología, la antropología, etc., ¿por qué no podemos decir que lo que entendemos por «traducción» es *traducción según los manuales que tienen esta propiedad?»* Putnam hace un diagnóstico exacto de la doctrina de Quine de la indeterminación especial de la traducción, diciendo que se deriva de una especie de esencialismo. Es, en términos generales, una doctrina en la que sabemos por adelantado que lo que no se puede formular en el vocabulario de la física de ese momento es tan poco esencial que está meramente «en el ojo del espectador», es sólo cuestión de conveniencia subjetiva. Véase también Christopher Boorse, «The Origins of the Indeterminacy Thesis», *Journal of Philosophy,* 72 (1975), 369-387, y Richard Rorty, «Indeterminacy of Translation and of Truth», *Synthese,* 23 (1972), 443-462.

nadas y de sus ontologías *es* significativo pero sólo con relación a la teoría básica con su propia ontología adoptada primitivamente y en el fondo inescrutable [28].

Podríamos pensar que esta tesis relativista es resultado natural y feliz del enfoque del conocimiento y de la ciencia que Quine comparte con Sellars, si no fuera por la inquietante expresión «ontología adoptada primitivamente y en el fondo inescrutable». En una concepción plenamente holística, la pregunta «¿Estamos refiriéndonos *realmente* a conejos o partes de un conejo?, ¿a fórmulas o a números Gödel?» no se consideraría ni como absurda ni como sensata únicamente cuando se pusiera en relación con un lenguaje básico [29], sino como igual a «¿Estamos hablando realmente de naciones o de grupos de personas individuales?» o «¿Estamos hablando realmente de brujas o de psicópatas con alucinaciones?». Las últimas preguntas tienen sentido si les damos sentido —es decir, si hay algo más que dependa de la respuesta. Es fácil imaginar situaciones en que se les daría sentido; es más difícil, si no imposible, en el caso del conejo-frente-a-las-partes-del-conejo. Pero a Quine no le interesa la cuestión de *esa* forma de dar sentido. Sus afirmaciones sobre la indeterminación y la inescrutabilidad no se suponen relacionadas con las necesidades de la ciencia o de la práctica. Admitir que a los lingüistas nunca se les ocurre sacar provecho de la indeterminación para traducir las expresiones cortas normalmente utilizadas cuando los conejos salen al descubierto en forma de «¡otra-parte-de-un-conejo!». Dice Quine:

> La máxima implícita que orienta su [del lingüista] elección de «conejo»... es que un objeto duradero y relativamente homogéneo, que se mueve como un todo sobre un fondo que contrasta, es una referencia probable a una expresión breve... La máxima es su propia imposición, en orden a establecer lo que es objetivamente indeterminado. Es una imposición muy sensata, y yo no recomendaría ninguna otra. Pero esto haciendo una afirmación filosófica [30].

[28] W. V. O. Quine, *Ontological Relativity and Other Essays* (Nueva York, 1969), págs. 50-51.

[29] Véase *ibíd.,* págs. 47 y ss. Hartry Field ha demostrado que la idea quineana de la «relativización a un lenguaje básico» y de «aceptar a pie juntillas la referencia» son incompatibles con su línea general de argumentación. Véase su «Quine and the Correspondence Theory», *Philosophical Review,* 83 (1974), páginas 207 y ss. Pero esta dificultad no tiene nada que ver con nuestro contexto.

[30] Quine, *Ontological Relativity,* pág. 34.

Una «afirmación filosófica» en este sentido es, cuando menos, una afirmación que no tiene ninguna importancia para determinar cómo es el mundo. Quine vacila entre la antigua concepción positivista de que estas afirmaciones son vergonzosamente «metafísicas» y la concepción más oxoniana de la filosofía-como-terapia de que estas afirmaciones peculiarmente filosóficas sirven como antídotos a un πρῶτον ψεῦδος como la «idea 'idea'». Sin embargo, podríamos considerar que esta afirmación filosófica concreta ofrece un antídoto contra las nociones de «ontología» y de «referencia». Es decir, podríamos volver a la concepción más anticuada de que lo mismo que el enfoque conductista de la «virtud en virtud del significado» en «Two Dogmas» no nos dejó ninguna noción de «igualdad de significado», salvo (como dice Harman) la de sentido común y sin interés filosófico en que «El presidente fue a Vietnam» y «Johnson fue a Vietnam» significan lo mismo, de la misma manera el enfoque conductista de la «ontología» de «Ontological Relativity» no nos deja ninguna noción de «igualdad de referencia», salvo la de sentido común y sin interés filosófico en que hablar de partes-de-un-conejo y hablar de conejos es hablar de la misma cosa (de formas diferentes)[31]. A Quine le parece que la noción filosófica de «referencia» contrasta con la de significado porque:

> La referencia, la extensión, ha sido lo firme; el significado, la intensión, lo no firme. Sin embargo, la indeterminación de la traducción con que nos enfrentamos ahora afecta a la extensión y a la intensión por igual. Los términos «conejo», «parte de conejo no separada» y «parte de un conejo» difieren no sólo en su significado; son verdaderos aplicados a cosas diferentes. La referencia misma resulta ser conductualmente inescrutable[32].

Pero esta firmeza relativa era mero producto de la afirmación de Quine de que las intensiones, para las que no había criterios de identidad, eran entidades más débiles que las extensiones, para las que sí los había. El problema de las condiciones de identidad para las intensiones se reduce, piensa Quine, al de cómo «deben relacionarse dos oraciones eternas para que, donde 'p' y 'q' hacen

[31] El ejemplo de la igualdad de significado basada en el sentido común procede de Harman, «Quine», pág. 142. Puede verse un mayor desarrollo de la distinción entre los sentidos filosófico y de sentido común de «hablar de» o referirse a, en mi «Realism and Reference», *Monist*, 59 (1976), 321-340, y capítulo 6, sección 4.

[32] Quine, *Ontological Relativity*, pág. 35.

sus veces, podamos decir que [p] es la misma proposición que[q] y no que otra»[33]. Pero pensar que se puede dar respuesta a esta pregunta es pensar, dice Quine, que hay alguna relación de sinonimia que hace una oración de un idioma la traducción correcta de una oración de otro idioma[34].

De esta manera hemos dado una vuelta completa. La firmeza de la referencia es lo que es debido a un supuesto contraste con una falta de firmeza sobre el significado. Pero esta falta de firmeza sólo se da si la traducción es indeterminada de alguna manera en que no lo es la física. Por eso, si aceptamos las críticas consagradas de la «doble» indeterminación de la traducción propuesta por Quine (indeterminación que difiere de la de la teoría física en que en el primer caso no hay ninguna «cuestión de hecho»), no tenemos ninguna razón para sorprendernos de que la referencia no consiga mejores resultados, ni ninguna razón para pensar que la inescrutabilidad conductual de la referencia lleva a ninguna conclusión, excepto «tanto peor para la referencia» o «partes de conejo y conejos son exactamente lo mismo». Como en este caso «referencia» significa una idea específicamente filosófica, cuya inescrutabilidad es una afirmación específicamente filosófica que depende de que se establezca entre conejos y partes de conejo una separación mayor de la que establecería cualquier necesidad científica o práctica, podríamos sentirnos con derecho a adoptar la misma actitud despreocupada hacia esta inescrutabilidad que adopta Quine hacia la idea específicamente filosófica de sinonimia y hacia la tesis de Brentano de la irreductibilidad de lo intencional.

En realidad, deberíamos adoptar esta actitud[35], pero no antes de examinar más de cerca la vacilación de Quine a propósito de la ontología. Decir que la idea filosófica de referencia es una idea de la que podríamos prescindir sin problemas es, como aceptaría Quine, decir que también lo es la ontología. Como es una preocupación por la ontología lo que hace a Quine tomar en serio la referencia, será útil ver qué difícil le resulta reconciliar esta preocupación con su afirmación holística de que no hay ninguna «filosofía primera» superior y anterior a la investigación científica corriente[36]. Esta última concepción debería inclinarle hacia la concepción de Sellars de que la «ciencia es la medida de todas las cosas,

---

[33] Quine, *Word and Object,* pág. 200.
[34] *Ibid.,* pág. 206.
[35] Expongo mis argumentos en contra en el capítulo 7.
[36] Véase «On Carnap's Views on Ontology» en Quine, *Ways of Paradox* (Nueva York, 1968) y «Epistemology Naturalized» en *Ontological Relativity.*

de qué es lo que es, y de qué no es lo que es»[37]. Sin embargo, Quine afirma que la imposibilidad práctica de prescindir de los giros intencionales no debería hacernos cerrar los ojos al hecho de que:

> Si estamos dibujando la estructura verdadera y última de la realidad, nuestro esquema canónico es el esquema austero que no conoce ninguna cita que no sea la cita directa ni actitudes proposicionales que no sean las constituciones físicas y conducta de los organismos[38].

Este proyecto, afirma, está en continuidad con el de la ciencia porque:

> Cada eliminación de construcciones o nociones oscuras que consigamos realizar, mediante paráfrasis en elementos más claros, es una clarificación del esquema conceptual de la ciencia. Los mismos motivos que impulsan a los científicos a buscar teorías cada vez más simples y claras adecuadas al objeto material de sus ciencias especiales son motivos de simplificación y clarificación del marco más amplio compartido por todas las ciencias... La búsqueda del patrón más simple y claro de notación canónica no debe considerarse distinta de una búsqueda de categorías definitivas, un dibujo de los rasgos más generales de la realidad. Y no se puede replicar que estas construcciones son medidas convencionales no dictadas por la realidad; pues, ¿no se puede decir lo mismo de una teoría física? Ciertamente, la naturaleza de la realidad es tal que una teoría física nos permitirá movernos mejor que otra; pero lo mismo ocurre con las notaciones canónicas (pág. 161).

El truco está, naturalmente, en saber qué son «oscuridad» y «claridad». Las *Geisteswissenschaften,* piensa Quine, utilizan nociones que son tan poco claras que lo único que podemos hacer al dibujar la estructura de la realidad es borrarlas. Todo es claro, sin embargo, en las ciencias físicas, excepto cuando recurren a números, funciones, propiedades, etc., en cuyo caso los interpretamos como conjuntos, interpretación que el científico físico puede considerar con sublime indiferencia. Pero la falta de claridad de «creencia», «significado», «se traduce por...», etc., es irremediable; no

---

[37] La frase está tomada de Sellars, *Science, Perception and Reality,* pág. 173

[38] Quine, *Word and Object,* pág. 221.

tenemos a mano nada que pueda reemplazarlos; pueden sobrevivir únicamente basándose en la comodidad práctica[39].

Pero, ¿por qué «cree en...» y «se traduce por...» deben a las necesidades de la práctica más que «es el mismo electrón que...» y «es el mismo conjunto que...? ¿Por qué las *Naturwissenschaften* dibujan la realidad mientras que las *Geisteswinssenschaften* simplemente nos permiten hacerle frente? ¿Qué es lo que las separa, dado que ya no pensamos en ninguna clase de afirmación que tenga condición epistemológica privilegiada, sino que todas las afirmaciones funcionan en unión para el bien de todos en ese proceso de adaptación holística gradual hecho famoso por «Two Dogmas of Empiricism»? ¿Por qué la unidad de investigación empírica no es toda la cultura (incluyendo las *Natur-* y las *Gesteswissens chaften)* y no sólo la totalidad de la ciencia física?

Si intentamos responder a estas cuestiones retóricas incurrimos en una auténtica contradicción, en opinión de Quine. Aparece con toda claridad en un pasaje en que está tratando de demostrar que los dictados prácticos de la traducción no tienen implicaciones epistemológicas:

> «Salvo la verdad lógica» tiene carácter convencional debido a la indeterminación de la traducción... La misma necesidad de determinación concede un valor especial a la adhesión a esta regla fuerte y simple en cuanto determinante parcial... «Salvo la verdad lógica» es una convención y una convención prudente. Vemos también que no da a la verdad lógica ninguna

[39] Harman («Quine», pág. 126) presenta una interpretación más comprensiva de las ideas de Quine en este terreno. Según Harman,

> no es que él crea que los objetos intensionales, las proposiciones o significados sean una entidad de carácter *extraño* (como podría creerse que los electrones son una realidad de carácter extraño). De lo que se queja no es de que los objetos intensionales, por ser algo abstracto, ofendan a sus sensibilidades en la forma en que, indudablemente, ofenden a las sensibilidades de Nelson Goodman... El argumento de Quine... es que las diversas concepciones de ese conjunto [que invocan tales entidades] son teorías que no explican lo que tratan de explicar. Por eso, su actitud hacia los objetos intensionales es semejante a la que tiene hacia el flogisto o el éter (o las brujas).

Parece que Quine aprueba la interpretación de Harman en *Words and Objections*, pág. 296. Pero creo que esta interpretación no se puede reconciliar con muchos de los argumentos presentados en *Word and Object* y en otras obras, aunque reconozco que representa la actitud que Quine *debería* haber adoptado.

situación epistemológica distinta de la de cualquier verdad obvia de tipo factual[40].

Pero si la convencionalidad depende de una indeterminación *especial* de traducción, podemos *no* decir, como hace Quine en un párrafo citado anteriormente, que la teoría física es algo «convencional no dictado por la realidad». Si la permanencia de la verdad lógica es meramente una regla práctica, y no una visión de la naturaleza de la realidad, si la teoría física *es* dicha visión, no puede ser también una regla práctica.

Para resumir sus vacilaciones, podríamos señalar que Quine quiere afirmar todo esto:

1. Existe eso que se llama ontología, que está gobernada por «escrúpulos sobre qué objetos se pueden suponer» y basada en una distinción entre «reificación irresponsable y su contrario»[41].
2. No hay ninguna categoría epistemológica especial que tenga ninguna oración aparte de su papel en el mantenimiento de ese «campo de fuerza» que es el conocimiento humano y cuyo objetivo es hacer frente a las irradiaciones sensoriales.
3. Por tanto, no existe eso que llaman contacto directo con los datos sensoriales o significados que otorgarían inviolabilidad a los informes en virtud de su correspondencia con la realidad, aparte de su función en el esquema general de la creencia.
4. Por eso, la epistemología y la ontología nunca se unen, pues nuestros escrúpulos sobre qué objetos debemos suponer no están dictados por nuestro contacto con los universales o los particulares.
5. Pero hay, sin embargo, una distinción entre aquellas partes del tejido de la creencia que expresan cuestiones de hecho y aquellas que no lo hacen, y la ontología nos garantiza que podamos detectar esta diferencia.

Para afirmar (5) al mismo tiempo que (1)-(4), Quine tiene que dar un sentido a la distinción entre «cuestión de hecho» y «convención» que no tenga ninguna relación con la distinción instrumentalista-fenomenalista más usual —la que se da entre aquello de lo que tenemos realmente conocimiento y lo que «postulamos» para enfrentarnos a los estímulos. La única forma de hacerlo, por

[40] *Words and Objections*. pág. 318.
[41] Quine, *Word and Object*, págs. 119-120.

lo que yo puedo ver, es sencillamente considerar las partículas elementales de la física contemporánea como cuestiones de hecho paradigmáticas y explicar que el sentido en que no hay cuestión de hecho sobre los significados o creencias es que se pueden decir cosas diferentes de lo que significa una oración o cree un hombre sin ninguna implicación para los movimientos de esas partículas. Esta táctica da carácter puramente estético a su preferencia de la física sobre la psicología, y, por tanto, a su preocupación por «la reificación irresponsable». Además, no puede funcionar. Otras posibles teorías bioquímicas (por ejemplo), al igual que otras posibles teorías psicológicas, serán compatibles con todos y sólo los mismos movimientos de las mismas partículas. A no ser que y mientras no haya una deducción auténtica de todas las afirmaciones nomológicas verdaderas derivadas de las leyes de la física (cosa que nadie esperaría en serio), no se podrá hacer ninguna queja de las intenciones que no se pueda hacer de la mitocondria [42].

[42] Dagfinn Føllesdal («Meaning and Experience» en *Mind and Language*, ed. S. Guttenplan, Oxford, 1975) sugiere una forma de interpretar la tesis de la indeterminación de Quine orientada a demostrar que «la posición de Quine es más interesante si su inclinación ontológica hacia el fisicismo se entiende como consecuencia de una inclinación epistemológica más fundamental hacia el empirismo» (pág. 33). Sugiere que

> ... todas las verdades existentes están incluidas en la teoría de la naturaleza. Como hemos señalado antes, en nuestra teoría de la naturaleza tratamos de explicar *todas* nuestras experiencias. Y las únicas realidades que podemos suponer justificadamente son aquellas a las que se recurre en la teoría más simple que explique todas estas pruebas. Estas entidades y sus propiedades e interrelaciones son todo lo que hay en el mundo, y todo aquello de lo que se puede tener razón o estar equivocado. Todas las verdades sobre ellas se incluyen en nuestra teoría de la naturaleza. En la traducción no estamos describiendo otro ámbito de la realidad, simplemente estamos correlacionando dos teorías globales que se refieren a todo lo que hay (página 32).

Sin embargo, no sé cómo podemos ver cuándo hemos dejado de describir y comenzado a correlacionar descripciones. O, dicho con otras palabras, no se me ocurre cómo podemos delimitar la «naturaleza» de alguna otra cosa, como no sea encontrando algún sentido de la expresión «explicar todas nuestras experiencias» en que algo que no llegue a ser toda la cultura pueda explicar todas nuestras experiencias.

Otra forma de remontar la tesis de la indeterminación hasta el empirismo nos la ofrece John McDowell («Truth Conditions, Bivalence, and Verificationism» en *Truth and Meaning*, ed. G. Evans y J. McDowell [Oxford, 1976]). McDowell piensa que Quine puede estar exponiendo «una versión, formulada sin mucha fortuna, de la objeción verificacionista fuerte al realismo en una

Quine se ve en estas dificultades, creo yo, por el intento de mantener la idea que él, igual que Sellars, hereda de Carnap y en último término del *Tractatus* de Wittgenstein: la idea de que el mundo pueda ser «descrito completamente» en un lenguaje extensional. El verdadero problema es la intensionalidad, más que la intencionalidad, pues sólo el carácter funcional de no verdad del discurso intencional hace que su supuesto objeto de estudio sea menos respetable que, por ejemplo, el hablar, en forma irreductiblemente bioquímica, de la mitocondria. La reductibilidad del hablar de partículas no es más que una tapadera de la reductibilidad al discurso con valor funcional de verdad. Las partículas no importan, pero la forma lógica sí. La falta de condiciones de identidad claras para las intenciones es un desastre no porque se derive cierto carácter fantasmagórico, sino sencillamente porque esta ausencia deja algunas oraciones no-extensionales. Pero si esto es así, podemos conseguir los fines de Quine sin utilizar sus medios. Podemos hacerlo admitiendo que el mundo *puede* describirse completamente en un lenguaje con valor funcional de verdad, y admitiendo al mismo tiempo que partes de él se pueden describir también en un lenguaje intensional, y absteniéndonos simplemente de hacer comparaciones odiosas entre estos modos de descripción. Decir que puede ser *completamente* descrito es utilizar una noción de completo en términos de extensión espacio-temporal, no en términos de capacidad explicativa ni de conveniencia práctica. Si no pudiéramos referirnos a las intenciones, nos sería difícil enfrentarnos con el mundo, pero aun así seríamos capaces de describir todas y cada una de sus partes, e incluso hacer una predicción acertada sobre el contenido de cualquier región espacio-temporal de la minuciosidad deseada.

---

teoría del significado» (pág. 65). Esta objeción es que interpretar la verdad de una afirmación como «infradeterminada por lo que es observable» nos obligaría, si interpretáramos la afirmación «realistamente», a atribuir al hablante «una concepción de la verdad como si fuera independiente de lo que es observable» (pág. 64). Como, para el verificacionista, esto último es absurdo, se demuestra que no debemos interpretar la afirmación «realistamente». No obstante, esta estrategia parece que supone encontrar un sentido a «determinación por lo que es observable» que mantuviera la biología y excluyera la traducción y, una vez más, no veo cómo se puede realizar esto. Por eso, concluyo que la tensión entre (4) y (5) sigue en pie a pesar de los intentos de los críticos comprensivos con Quine por reformular su afirmación de forma que resulte a salvo de la crítica de Chomsky de que la única indeterminación en el área es la conocida infradeterminación de la teoría por la observación (crítica que mencionan Follesdal y McDowell, y que quieren evitar).

La forma de aplicar este punto al vocabulario de las creencias y deseos nos la indica Davidson, quien expone la cuestión en términos de una distinción entre generalizaciones homonómicas y heteronómicas:

> Por una parte, hay generalizaciones cuyas instancias positivas nos dan razones para creer que se podría mejorar la generalización misma introduciendo nuevas salvedades y condiciones formuladas en el mismo vocabulario general que la generalización original. Esta generalización apunta a la forma y vocabulario de la ley terminada; podemos decir que es una generalización *homonómica*. Por otra parte, hay generalizaciones que cuando se manifiestan en casos concretos pueden darnos motivos para creer que hay en vigor una ley precisa, pero una ley que sólo se puede formular cambiando a un vocabulario diferente. Podemos llamar a estas generalizaciones *heteronómicas*.
>
> Supongo que la mayor parte de nuestro saber práctico (y de la ciencia) es heteronómica. Esto se debe a que una ley sólo puede esperar ser precisa, explícita y con las menores excepciones posibles si extrae sus conceptos de una teoría cerrada general... La confianza en que una formulación es homonómica, corregible dentro de su propio dominio conceptual, exige que extraiga sus conceptos de una teoría con fuertes elementos constitutivos...
>
> Lo mismo que no podemos asignar inteligiblemente una longitud a ningún objeto a no ser que haya una teoría general que valga para los objetos de esa clase, tampoco podemos atribuir inteligiblemente ninguna actitud proposicional a un agente a no ser dentro del marco de una teoría viable de sus creencias, deseos, intenciones y decisiones [43].

Davidson continúa diciendo que las supuestas leyes psicofísicas son como «Todas las esmeraldas son 'verzules'». Combinan términos tomadas de vocabularios dispares. Podemos hablar de esmerrosas y de verzuleidad, o de esmeraldas y verdor, pero no de ambas cosas a la vez (al menos no si queremos una teoría general útil). Aún así, podemos hablar de acciones y creencias, o de movimientos y

[43] Donal Davison, «Mental Events» en *Experience and Theory*, ed. L. Foster y J. W. Swanson (Amherst, Mass., 1970), págs. 94-96.

N. *del T.*—«Grue» es una palabra artificial en la que se unen elementos de «GReen» (=verde) y «blUE» (=azul). Por eso, utilizamos la palabra «verzul», compuesta con elementos de «VERde» y «aZUL». Lo mismo ocurre con «emeroses»=EMERald (=esmeralda)+«rOSE» (=rosa), que traducimos por «esmerrosa».

neuronas, pero no (globalmente) de ambas cosas a la vez. Pero hay un sentido obvio, en el primer ejemplo, en el que estamos hablando de *las mismas cosas,* cualquiera que sea el conjunto de predicados que elijamos. E incluso, dice Davidson, en el segundo caso. La diferencia en la elección de vocabularios no es una señal distintiva de la diferencia entre lo real y lo ontológicamente desprestigiado, no de la que existe entre lo factual y lo mítico, sino que es semejante a la diferencia entre hablar de las actividades de los ministros y generales, o entre hablar de la mitocondria en cuanto tal y hablar de las partículas elementales que contienen. Podemos decir verdadera y prácticamente cosas como «Si Asquith hubiera seguido de primer ministro, Inglaterra habría perdido» o «Si hubiera habido algunos neutrones más ahí, no habría sobrevivido el mitocondrio», o «Si hubiéramos colocado un electrodo en el lugar adecuado de la corteza, nunca habría llegado a la conclusión de que era Napoleón», o «si pudiéramos hacernos con una esmerrosa, tendríamos el matiz de verdad adecuado», pero no podemos (que nosotros sepamos) convertir estas observaciones heteronómicas en leyes que formen parte de teorías generales. Y, por otra parte, tampoco tenemos que pensar que estas observaciones heteronómicas atraviesan la línea que divide las esferas ontológicas —en especial, la esfera de lo factual y la esfera de lo no factual. En la concepción de Davidson de la relación entre diferentes vocabularios explicativos, no hay ninguna razón *en absoluto* para pensar que los vocabularios que se prestan a formulaciones con valor funcional de verdad «retratan la estructura verdadera y última de la realidad» de una manera en que no pueden hacerlo los vocabularios intensionales. La distinción extensional-intensional resulta que no tiene ni más ni menos interés filosófico que la distinción entre naciones y personas: es capaz de incitar la emoción reduccionista, pero no de ofrecer una razón especial para embarcarse en proyectos reduccionistas.

La distinción de Davidson nos ofrece una forma de ver que un vocabulario intencional no es más que un vocabulario más para hablar de las partes de un mundo que puede describirse, realmente, sin este vocabulario. Podemos compartir la intuición de Carnap de que el movimiento de cualquier cosa se puede predecir tomando como base el movimiento de las partículas elementales, y que si estuviéramos tomando nota de todas las partículas, estaríamos tomando nota de (aunque *no* explicando) todo lo que existe, sin hablar, con Quine, de «lo infundado de los giros intencionales y del vacío de una ciencia de la intención». Un vocabulario —el de la física de partículas— puede servir para *todas y cada una* de las

partes del universo, mientras el hablar de mitocondria, esmerrosas, ministros del gobierno e intenciones es algo que sólo se requiere en algún que otro lugar. Pero la distinción entre lo universal y lo específico no es la distinción entre lo factual y lo «vacío», y todavía menos la que hay entre lo real y lo aparente, o lo teórico y lo práctico, o la naturaleza y la convención.

No obstante, Davidson vincula su propio proyecto al de Quine de forma engañosa cuando dice que «el carácter heteronómico de las afirmaciones generales que unen lo mental y lo físico se remonta a esta función central de la traducción en la descripción de todas las actitudes proposicionales, y a la indeterminación de la traducción» [44], y también cuando cita aprobatoriamente la observación de Quine de que «la tesis de Brentano de la irreductibilidad de los giros intencionales es parecida a la tesis de la indeterminación de la traducción» [45]. Ambas afirmaciones dan a entender que la relación entre las formulaciones que ofrecen traducciones y el lenguaje de la conducta es *especial* en una forma en que no lo es la relación entre afirmaciones sobre la mitocondria y las afirmaciones sobre las partículas elementales. Ambas sugieren la extraña doctrina de Quine de la «doble» indeterminación de la traducción. Pero si tengo razón en lo que vengo diciendo, la irreductibilidad es siempre *sólo* irreductibilidad, y nunca una pista de las diferencias «ontológicas». Hay *muchos* vocabularios en el lenguaje dentro de los cuales podría esperarse conseguir una teoría general expresada en generalizaciones homonómicas, y la ciencia, la teoría política, la crítica literaria, etcétera, seguirán, Dios mediante, creando cada vez nuevos vocabularios de este tipo. Abandonar la idea de que en la filosofía tenemos una disciplina que protege de la «reificación irresponsable» y sistematiza nuestros «escrúpulos sobre qué objetos podemos suponer» sería acomodar la irreductibilidad a nuestras necesidades y, por tanto, juzgar cada uno de estos vocabularios en términos pragmáticos o estéticos únicamente. Las críticas de Quine al intento de Carnap de separar la filosofía de la ciencia son precisamente lo que hace falta para ayudarnos a comprender que no existe tal disciplina, y, por tanto, lo que hace falta precisamente para ver que la *Geisteswissenschaften* no habría devenido más *wissenschaftlich*, o más ontológicamente respetable, si hubiera resultado que Brentano y Dilthey estaban equivocados al hablar de su irreductibilidad. Pero por desgracia, la firme convicción de Quine de que la lógica simbólica debe tener de alguna manera «implicaciones ontológicas» le

[44] *Ibid.*, pág. 97.
[45] *Ibid.*, pág. 97n., donde cita a Quine, *Word and Object*, pág. 221.

impulsa a dar más importancia de la necesaria a la traducción, a la intencionalidad y a la «idea 'idea'».

He dedicado esta larga sección a demostrar que el ataque de Quine a la «verdad en virtud del significado» en cuanto explicación de las verdades supuestamente necesarias no debe confundirse con su ataque a los «significados» en cuanto ideas que se dan en la mente, ideas que determinan la exactitud de las traducciones en una forma que a la conudcta lingüística le resulta imposible. La primera es realmente una seudo-explicación; no hay, por las razones holísticas presentadas en «Two Dogmas», representaciones *privilegiadas*. Pero la desconfianza de Quine ante las representaciones privilegiadas le lleva a desconfiar de *todas* las representaciones, a desconfiar de la misma «idea 'idea'». Sin embargo, las ideas de la mente no son ni más ni menos dignas de crédito que las neuronas del cerebro, la mitocondria de las células, las pasiones del alma, o el progreso moral de la historia. El daño hecho por la «idea 'idea'» en la filosofía moderna fue producido por la seudo-explicación de la autoridad epistémica a través de la noción de «conocimiento directo» por el «Ojo de la Mente» de entidades mentales como los datos sensoriales y los significados. Pero éste es un daño epistemológico, no ontológico. Si tengo razón en las críticas a Quine que vengo presentando (y en la línea general que sigo en el libro), la única forma de producir daño *ontológico* es bloquear el camino de la investigación insistiendo en una toría antigua deficiente a expensas de una nueva teoría satisfactoria. Puede decirse que la psciología introspectiva del siglo XIX, bloqueó brevemente el camino de la investigación, pero aunque esto fuera así sería muy distinto de decir que las *Geisteswisenschaften* nos impiden ver claramente la realidad, o que haya que tolerar su ontología sospechosa por motivos prácticos. La lección del conductismo epistemológico es precisamente que no hay ninguna «afirmación filosófica» que hacer sobre la traducción o la intencionalidad, no sobre cualquier otro tema «ontológico». Más bien, nos ayuda a ver que la capacidad explicativa está donde la encontramos, y que el intento filosófico de distinguir entre explicaciones «científicas» y «no científicas» resulta innecesario.

5. Conductismo epistemológico, conductismo fisiológico y lenguaje

En el capítulo anterior he indicado que la tradición epistemológica confundía el proceso causal de adquirir el conocimiento con

las cuestiones relacionadas con su justificación. En este capítulo he presentado la crítica de Sellars al Mito de lo Dado y la crítica de Quine a la idea de verdad en virtud del significado, como dos manifestaciones concretas de esta crítica más general. Si aceptamos estas críticas, y, por consiguiente, abandonamos la idea de la epistemología como búsqueda, iniciada por Descartes, de los aspectos privilegiados en el campo de la conciencia que son piedras de toque de la verdad, estamos en situación de preguntar si queda todavía algo que pueda ser la epistemología. Quiero demostrar que no. Para entender las materias que Descartes quería entender —la superioridad de la Nueva Ciencia sobre Aristóteles, las relaciones entre esta ciencia y las matemáticas, el sentido común, la teología y la moralidad— hemos de dirigirnos hacia afuera en vez de hacia adentro, hacia el contexto social de la justificación más que a las relaciones entre las representaciones internas. Esta actitud ha sido fomentada en las últimas décadas por muchos progresos filosóficos, especialmente los procedentes de las *Philosophical Investigations* de Wittgenstein y la *Structure of Scientific Revolutions* de Kuhn. Algunos de estos progresos serán examinados a fondo en los capítulos siete y ocho. Pero antes de llegar a ese punto, quiero ocuparme de dos intentos de mantener algo de la tradición cartesiana, intentos que pueden parecer arrojar alguna duda sobre nuestra capacidad de abandonar sin más la imagen del Espejo de la Naturaleza.

El primero de estos intentos es la rebelión contra el conductismo lógico en la filosofía de la psicología, que lleva al desarrollo de las explicaciones de la conducta en términos de representaciones internas sin que haya, necesariamente, ninguna vinculación con la justificación de las creencias y acciones. Ya he dicho que una vez que se mantienen aparte la explicación y la justificación no hay ninguna razón para objetar a la explicación de la adquisición del conocimiento en términos de representaciones, y que estas explicaciones pueden presentarse sin resucitar el tradicional «problema mente-cuerpo». Pero creo que la defensa de estas explicaciones frente a Ryle y a Skinner se puede deformar fácilmente hasta convertirse en una rehabilitación de la tradicional problemática filosófica del siglo XVII, y por eso dedicaré el capítulo cinco a examinar estas defensas. Mi intensión será desligar la psicología empírica de los restos de la epistemología defendiéndola frente a las críticas wittgensteinianas y los piropos chomskianos.

El segundo intento de conservar algo de la tradición cartesiana del que me voy a ocupar es el esfuerzo, dentro de la actual filosofía del lenguaje, por especificar «cómo conecta el lenguaje con

el mundo», creando así una analogía del problema cartesiano de cómo conecta el pensamiento con el mundo. Cualquier intento de utilizar nociones de referencia de términos y verdad de las oraciones para ayudar a comprender las cuestiones que preocuparon a Descartes me parece condenado al fracaso, pero se trata de un programa muy tentador. Dado que el lenguaje es un Espejo de la Naturaleza «público», así como el pensamiento lo es «privado», parece que deberíamos ser capaces de reformular muchísimas cuestiones y respuestas cartesianas y kantianas en términos lingüísticos, y, por tanto, rehabilitar una gran parte de las cuestiones filosóficas clásicas (por ejemplo, la elección entre idealismo y realismo). El capítulo seis está destinado a tratar varios esfuerzos por conseguir dicha rehabilitación y a demostrar que la semántica debe mantenerse tan libre de la epistemología como la psicología.

En el momento en que consideramos irrelevantes para los problemas de la justificación las representaciones internas necesarias en la explicación psicológica y las relaciones palabra-mundo necesitadas por la semántica para producir una teoría del significado para los lenguajes naturales, podemos interpretar el abandono de la búsqueda de representaciones privilegiadas como el abandono de la aspiración a una «teoría del conocimiento». El deseo del siglo XVII por llegar a esta teoría fue producto del cambio de un paradigma de comprensión de la naturaleza a otro, así como del cambio de una cultura religiosa a otra secular. La filosofía como disciplina capaz de ofrecernos un «método correcto de buscar la verdad» depende de que se encuentre algún marco neutro y permanente de todas las posibles investigaciones, cuya comprensión nos permitirá ver, por ejemplo, por qué ni Aristóteles ni Belarmino tenían razón para creer lo que creían. La mente en cuanto Espejo de la Naturaleza fue la respuesta de la tradición cartesiana a la necesidad de este marco. Si no hay representaciones privilegiadas en este espejo, ya no dará respuesta a la necesidad de una piedra de toque para elegir entre llamadas justificadas e injustificadas a nuestra creencia. Si no se encuentra otro marco de esta naturaleza, el abandono de la imagen del Espejo nos lleva al abandono de la idea de la filosofía como disciplina que juzga las afirmaciones de la ciencia y la religión, de las matemáticas y la poesía, de la razón y el sentimiento, atribuyendo a cada una su lugar adecuado. Desarrollaré más detalladamente este punto en los capítulos siete y ocho.

Capítulo V

Epistemología y Psicología empírica

1. Sospechas sobre la Psicología

La línea de pensamiento que he llamado «conductismo epistemológico» ha producido un prejuicio contra la idea de «entidades mentales» y «procesos psicológicos». Una cierta imagen de las facultades superiores del hombre, común a Descartes y Locke, se ha ido borrando poco a poco debido a las obras de autores como Dewey, Ryle, Austin, Wittgenstein, Sellars y Quine. Sin embargo, esta imagen —la que dio lugar a la idea del «velo de ideas» en el siglo xvii, y, por tanto, al escepticismo epistemológico— no ha sido reemplazada por otra imagen nueva más clara. Por el contrario, entre los anti-cartesianos se observa un desacuerdo fratricida sobre qué hay que decir, si se debe decir algo, sobre la mente. La palabra mágica de Ryle, «disposición», ya no está de moda, y se presentan como sustitutos ideas más actualizadas, como la de «estado funcional». Se mira con recelo todo lo que huela a conductismo metodológico skinneriano o a conductismo «lógico» ryleano, pero se está de acuerdo en que debe haber alguna forma de evitar estos esfuerzos reductistas sin recaer en el tipo de dualismo que originó los tradicionales «problemas de la filosofía moderna». Las consecuencias contra-intuitivas del conductismo, interpretado reductivamente, están ilustradas en la polémica de Malcolm contra gran parte de las obras recientes de los psicólogos:

> Así pues, son los hechos, las circunstancias que rodean a la conducta, quienes le dan la propiedad de expresar reconocimiento. Esta propiedad no se debe a algo que ocurra en el interior.
> Me parece que si los filósofos y psicólogos entendieran esto, no tendrían ya motivo para construir teorías y modelos para el

reconocimiento, la memoria, el pensamiento, la resolución de problemas, la comprensión y otros «procesos cognitivos» [1].

Si seguimos esta línea de pensamiento, podemos llegar a la conclusión de que toda la ciencia empírica de la psicología está basada en un error, y de que no queda un campo intermedio de investigación entre las explicaciones de la conducta basadas en el sentido común, por una parte, y la neurofisiología, por la otra. Según esta concepción, la idea de que la psicología cuenta con un terreno intermedio de investigación será producto del «mito de las estructuras y procesos cognitivos» de Malcolm —de lo que Ryle llamaba «Mito de Descartes». Malcolm parece inclinado a llevar las cosas a este punto, por ejemplo, en su descripción de la idea chomskiana de un «sistema interiorizado de reglas» como característico del error radical de la «teoría tradicional de las Ideas», a saber:

> La suposición... de que al hablar una persona ha de estar *guiada*. Debe tener a mano algo que le indique cómo ha de hablar, cómo unir las palabras gramaticalmente y con un sentido coherente... Lo que se está explicando es el conocimiento —tanto el conocer como el saber cómo. Se supone que la presencia en él de la estructura del lenguaje o de su sistema de reglas explica este conocimiento —que explica cómo sabe (página 389).

Si llegamos a ver que «nuestra comprensión de las capacidades cognitivas humanas no progresa por el hecho de sustituir la mitología del estímulo-respuesta por una mitología de sistemas de orientación interior» (pág. 392), piensa Malcolm, no pensaremos que haya ninguna explicación que se pueda buscar en este campo.

Es totalmente cierto que el modelo de la mente que llevó a Descartes y a Locke a construir los «problemas filosóficos tradicionales» se incorporó a la terminología de la joven ciencia de la psicología [2]. Sería sorprendente que el derrocamiento de este modelo no repercutiera en las actividades de dicha ciencia. Sin embargo, sería igualmente sorprendente que una disciplina, que ha visto pasar varias generaciones desde el momento de sus inicios filosóficos, no fuera capaz de valerse por sí misma. Tiene que haber, sospechamos nos-

---

[1] Norman Malcolm, «The Myth of Cognitive Processes and Structures», en *Cognitive Development and Epistemology,* ed. Theodore Mischel (Nueva York y Londres, 1971), pág. 387.

[2] Véase, por ejemplo, J. C. Flugel y Donald J. West, *A Hundred Years of Psychology* (Londres, 1964). capítulos 1 y 2.

otros, programas de investigación psicológica que no puedan verse en peligro ante la crítica filosófica de los vocabularios utilizados por sus diseñadores.

Las críticas wittgesnteinianas (como las de Malcolm) a estos programas parecen muchas veces basadas en el paso erróneo de

1. El significado de los términos que hacen referencia a lo mental debe explicarse en términos de conducta (donde «conducta» es una abreviatura de «funciones que relacionan circunstancias y estímulos con la conducta») y no por ostensión interior

a

2. La psicología sólo puede ocuparse de las correlaciones empíricas entre actos de conducta y circunstancias externas.

La inferencia, como han señalado varios críticos de Ryle, no tiene mayor fuerza que la inferencia paralela hecha por los filósofos operacionalistas de la ciencia sobre la física[3]: «Fodor, por ejemplo, señala que el psicólogo está perfectamente preparado para admitir que ciertos rasgos de la conducta y de la matriz social son condiciones necesarias para la ocurrencia de pensamientos, reconocimientos, emociones, etc., pero insiste en que puede haber también gran cantidad de condiciones 'internas' igualmente necesarias»[4]. Por eso, si el investigador psicológico tiene sensatez suficiente para evitar *definir,* por ejemplo, un «acto de reconocimiento» en términos únicamente de hechos internos, probablemente podrá sacar partido de la conducta y entorno para identificar sus datos. ¿Qué más que esto se puede necesitar para evitar la acusación de mitologizar? Podría ocurrir, naturalmente, que no hubiera «variables intervinientes», pero es de suponer que esto sólo pueda descubrirse a posteriori, mediante tanteos. Como dice P. C. Dodwel, en réplica a Malcolm:

> Según la concepción de Malcolm, los psicólogos tendrían que limitarse a investigaciones de simples relaciones empíricas como la que podría obtenerse entre memoria y privación del sueño. Pero es enormemente difícil ver cómo se puede justificar esta restricción. Los factores que investigan los psicólogos en la

[3] Creo que el primero en afirmarlo fue Albert Hofstader, en «Professor Ryle's Category-Mistakes», *Journal of Philosophy,* 47 (1951), 257-270.

[4] Jerry Fodor, «Could There Be a Theory of Perception?», *Journal of Philosophy,* 63 (1966), 371.

memoria humana *son* relaciones empíricas, aunque generalmente de mayor complicación que la que acabamos de mencionar. ¿Quién es, entonces, el que ha de decir qué relaciones empíricas deben estar abiertas a la investigación? Los filósofos, no, por supuesto [5].

Esta respuesta no me parece totalmente convincente, pero, sin embargo, podemos ganar algo considerando las razones por las que el operacionalismo reductivo ha parecido mucho más aceptable para la psicología que para la física. ¿Por qué han regateado los filósofos a los psicólogos el derecho a idear cualquier entidad y proceso teórico que pudiera llevarles a explicar nuestra conducta? Ya hemos expuesto una de las razones: la confusión entre las afirmaciones que hemos señalado anteriormente con el (1) y con el (2). Esta confusión se basa en el temor, existente en Husserl y en Dilthey, así como en wittgesnteinianos como Winch y Kenny, de que el someter la conducta humana a una explicación mecanicista en términos de «procesos psicológicos» oscurecería la distinción entre personas y cosas, entre la realidad humana estudiada por las *Geisteswissenschaften* y el resto de la realidad estudiado por las *Naturwissenschaften*. En los capítulos siguientes examinaremos esta distinción, pero por ahora nos contentaremos con la respuesta de Dodwell. El miedo al mecanismo y a la pérdida de personalidad es un motivo de sospecha de *todas* las ciencias conductuales, y no capta lo que los filósofos han considerado especialmente dudoso en relación con la psicología. Las razones de sospecha más directamente relevantes son aquellas que indican que los psicólogos deberían ser *más* mecanicistas, y no menos, que deberían atravesar directamente lo mental y llegar hasta lo neurofisiológico.

La primera razón es simplemente el impulso hacia la ciencia unificada, que no es tanto el impulso a reducir lo Plural a lo Uno cuanto la convicción de que la ciencia del siglo XVII descubrió que todo puede explicarse mediante átomos y vacío, y que la filosofía tiene el deber moral de conservar esta visión. Sin embargo, esta convicción se ha visto debilitada por una oscura conciencia de la mecánica cuántica, por lo que el respeto ontológico a la materia no sensible ha sido sustituido por un respeto sociológico hacia los profesores de física. Las referencias de los filósofos a «lo físico» van ahora acompañadas generalmente de una nota que explica que

[5] P. C. Dodwell, «Is a Theory of Conceptual Development Necessary?», en *Cognitive Development and Epistemology* (citado en la nota 1, más arriba), pág. 382.

se interpretará como «física» cualquier realidad que sea mencionada por las «ciencias físicas». Antes de Quine, cuando la «reducción» seguía siendo el núcleo del programa lógico-empírico, los filósofos se consideraban capaces de realizar aportaciones a la unificación de la ciencia mediante el «análisis de significados» de términos utilizados en sociología, psicología, etc. Pero a partir del ataque de Quine al significado la necesidad de reducir todo a lo-que-aprueben-los-físicos se ha visto reemplazada por una sensación todavía más vaga de que las ciencias que no son la física se tornan «más científicas» cuando pueden sustituir las descripciones funcionales de las entidades teóricas (por ejemplo, «gene») por descripciones estructurales (por ejemplo, «molécula ADN»). Esta sensación es efímera en casos como la sociología y la economía, donde nadie tiene interés en insistir en la realización física de las realidades teóricas postuladas, pero persiste en la psicología, cuyas entidades teóricas tienen más o menos el tipo de concreción que invita a la sustitución por la neurofisiología. Pero, dadas las razones quineanas para pensar que no se pueden presentar en términos recíprocos las condiciones interesantes necesarias y suficientes para la aplicación de los términos de una disciplina, ¿debemos estar impacientes por que se produzca dicho reemplazo? Nadie pensó que la genética implicara recurrir a entidades dudosas porque el ADN tardara mucho tiempo en llegar. Así pues, ¿qué explica la sensación instintiva de que los psicólogos están bloqueando el camino de la investigación?

Para responder a esto, debemos considerar una segunda razón de sospecha ante los procesos y entidades mentales postulados. Siguiendo a Ryle, la llamaremos el miedo a los espíritus. El pensamiento de que apoyando lo mental, incluso temporalmente, estamos perdiendo espíritu científico, proviene de dos fuentes. La primera, que ya he mencionado en el capítulo primero, es la confusión de la concepción post-cartesiana de la «conciencia» con la noción prefilosófica del alma como aquella que abandona al cuerpo en el momento de la muerte. La segunda es el argumento epistemológico de que la introspectibilidad lleva consigo el acceso privilegiado, y que dado que este privilegio epistémico ha de estar basado en una diferencia ontológica (las entidades mentales son intrínsecamente mejor conocidas a sus posesores de lo que nadie puede conocer algo físico), hemos de negar la existencia de lo mental-en-cuanto-introspectible so pena de hacer que parte de nuestro conocimiento de la realidad dependa de informes no verificables. Este argumento no suele presentarse en forma tan descarada, pero algo parecido a esto es lo que subyace en gran parte de la hostilidad positivista y wittgenstei-

niana contra lo mental[6]. Sin embargo, lo mismo que la noción de la «unidad de la ciencia» en cuanto programa de investigación filosófica no puede sobrevivir el ataque de Quine contra el «significado», tampoco este argumento epistemológico puede sobrevivir al tratamiento que Sellars hace de «lo dado». En la explicación del conocimiento inmediato que hace Sellars, la introspección es una capacidad aprendida, y tiene muy poca justificación la oscura sospecha de que los sujetos llegarán a hacer introspección de todo lo que el experimentador les diga que deben ser capaces de someter a introspección. Según Sellars, nuestro conocimiento de los hechos mentales no es señal de una categoría ontológica distinta, y la incorregibilidad de los informes en primera persona, como todas las cuestiones relacionadas con la condición epistémica, es una cuestión sociológica, más que metafísica. Pero el renunciar a la condición fantasmagórica especial que se suponía hacía posible el acceso privilegiado libera a la respetabilidad metodológica de recurrir a la introspección. Ahora podemos ver cómo el enseñar a las personas a realizar introspección de los pensamientos, nostalgias, presión sanguínea o ritmos alfa extraños, es sencillamente cuestión de utilizar conexiones existentes dentro del organismo —probablemente las conexiones entre el centro del habla y el resto del sistema nervioso— como instrumentos científicos. El hecho de que este adiestramiento deba tener como punto de partida las circunstancias disponibles intersubjetivamente es suficiente para asegurar que no está pasando nada malo. Así pues, el carácter «acientífico» y la «subjetividad» de los informes introspectivos no son más significativos, filosóficamente hablando, que los defectos de los espectroscopios. Una vez que los «informes subjetivos» se consideran una cuestión de conveniencia heurística, en vez de permitir que la palabra no confirmada de alguien refute una hipótesis científica prometedora, podemos eliminar las desafortunadas asociaciones de la psicología introspectiva con la llamada del racionalismo a las ideas claras y distintas y la del protestantismo a la conciencia individual.

Mi conclusión es que los argumentos quineanos y sellarsianos que he esbozado en el capítulo anterior sirven también para librar a la psicología de las sospechas clásicas de los filósofos empiristas y fisicistas contra ella. Las sospechas que proceden de la otra dirección —de la necesidad de conservar la singularidad humana, la libre voluntad y la integridad de las *Geisteswissenschaften*— las estudia-

---

[6] Véase, por ejemplo, Malcolm y su «teoría de la expresión» de los informes mentales en primera persona, basada en las *Philosophical Investigations*.

remos en los capítulos séptimo y octavo. En el presente capítulo, me limitaré a la pregunta: ¿Podemos ver alguna relevancia de los problemas filosóficos tradicionales relacionados con el *conocimiento* en el sentido de que mejoren los resultados actuales o posibles de la investigación psicológica empírica? Como lo que voy a decir es que estos «problemas filosóficos» deberían suprimirse más que resolverse, es previsible que mi respuesta va a ser negativa. Pero esta respuesta negativa necesita una defensa importante, pues muchos filósofos que se han dejado impresionar por los argumentos contra las representaciones privilegiadas, dirigidos por Quine y Sellars, quieren, sin embargo, reemplazar la epistemología «fundacionalista» tradicional por el uso de los resultados psicológicos para llegar a una teoría general de las representaciones internas. Mi argumento será que esta «nueva epistemología» no puede ofrecer nada que sea relevante a los temas de la justificación, y que, por tanto, es totalmente irrelevante para las exigencias culturales que llevaron a la aparición de la epistemología en los siglos XVII y XVIII. Y, por consiguiente, tampoco puede ayudar a mantener la imagen de la filosofía como disciplina que se mantiene aparte de la investigación empírica y explica la relevancia de los resultados de dicha investigación para el resto de la cultura.

Para demostrarlo, estudiaré dos sugerencias muy repetidas en las obras filosóficas más recientes, que atribuyen a la psicología un grado de importancia filosófica mayor del que se merece. La primera es la indicación de Quine de que la psicología puede investigar las «relaciones entre teoría y evidencia» que solían constituir el objeto material de la epistemología. En la sección 2 argumentaré que estas relaciones no pueden reformularse en términos psicológicos. La segunda es la afirmación de que las analogías entre estados de programa de los ordenadores y estados psicológicos de las personas y entre estados de «hardware» de los ordenadores y estados neurofisiológicos de los cuerpos dan un sentido nuevo e interesante a la idea de que nuestro conocimiento consiste en una «representación interna» del mundo. Esta afirmación ha sido desarrollada con mayor detalle por Fodor; en las secciones 3 y 4 trataré de demostrar que Fodor une un sentido de la «representación» en que las representaciones se pueden juzgar como exactas o inexactas, y otro en que eso no es posible. Estos dos sentidos, en mi opinión, marcan los dominios respectivos de la epistemología y la psicología.

2. La falta de naturalidad de la epistemología

En un ensayo titulado «Epistemology Naturalized», Quine estudia los diferentes problemas con que se han de encontrar los esfuerzos por conseguir un «fundamento para la ciencia» y finalmente considera la sardónica actitud de Wittgenstein hacia esta empresa:

> Carnap y los otros positivistas lógicos del Círculo de Viena habían conseguido ya dar al término «metafísica» un sentido peyorativo, como si fuera equivalente a carencia de significado; y a continuación vino el término «epistemología». Wittgenstein y sus seguidores, principalmente en Oxford, encontraron una vocación filosófica residual en la terapia: en curar a los filósofos de sus alucinaciones de que había problemas epistemológicos.
>
> Pero creo que en este momento puede ser más útil decir que la epistemología todavía se mantiene, aunque en un marco nuevo y con una categoría más clara. La epistemología, o algo semejante, encaja perfectamente como un capítulo más de la psicología y, por tanto, de la ciencia natural. Estudia un fenómeno natural, a saber, un sujeto humano físico. A este sujeto humano se le reconoce una cierta entrada experimentalmente controlada —ciertos patrones de irradiación de varias frecuencias, por ejemplo— y en la plenitud de los tiempos el sujeto presenta como salida una descripción del mundo externo tridimensional y de su historia. La relación entre la exigua entrada y la salida torrencial es una relación que nos sentimos impulsados a estudiar por la misma razón que siempre estimuló el estudio de la epistemología; es decir, para ver cómo se relaciona la evidencia con la teoría, y de qué maneras la propia teoría de la naturaleza transciende cualquier evidencia disponible[7].

Consideremos en primer lugar la afirmación de Quine de que el motivo que ha impulsado a la epistemología ha sido siempre «ver cómo se relaciona la evidencia con la teoría, y de qué maneras la propia teoría de la naturaleza transciende cualquier evidencia disponible». La mayoría de los historiadores intelectuales se han sorprendido de que lo que llamamos ahora «teoría del conocimiento» desempeñara un papel tan insignificante en el pensamiento de los escritores de antes del siglo XVII. Tal como entiende Quine la epis-

[7] W. V. O. Quine, *Ontological Relativity and Other Essays* (Nueva York, 1969), págs. 82-83.

temología, no es fácil ver por qué las cosas fueron así. Podríamos apuntar que en tiempo de Galileo y Descartes se intensificó la necesidad de optar entre teorías radicalmente distintas sobre los planetas y misiles balísticos, y que, por consiguiente, la mente de Occidente se vio sorprendida nuevamente por la forma en que «la propia teoría de la naturaleza transciende cualquier evidencia disponible». Pero esta sugerencia no tiene mucha consistencia. En la época antigua y medieval hubo muchas teorías encontradas sobre los cielos, pero hemos de hurgar un tanto para encontrar algo que se llame «epistemología» en Platón y Aristóteles, si es que la epistemología significa observar las diferencias entre teoría y evidencia y comparar las formas que permitan superarlas. Quizá prorrumpamos en gritos de alegría en algunos párrafos de *Teeteto* y de *De Anima;* los historiadores neo-kantianos de la filosofía griega, como Zeller, lo hacen a menudo. Pero nada relacionado con el tema aparecerá en ninguna otra parte en el siglo IV, y prácticamente nada en la *Analítica posterior* (donde Aristóteles, que estaba perfectamente informaod de la controversia entre las teorías científicas enfrentadas, y que competía con la mejor de ellas, estudia la categoría y metodología de la ciencia). Cuando, en el siglo XVII, el cartesianismo irrumpió en un mundo sobresaltado, no estaba proponiendo una nueva concepción sobre cuestiones debatidas desde hacía tiempo sobre las relaciones entre teoría y evidencia. Lo que hizo fue tomarse en serio preguntas que, como observara Gilson con cierta indignación, los escolásticos no habían querido formularse [8].

Para entender por qué el siglo XVII se interesó por la relación entre teoría y evidencia, hemos de preguntarnos por qué la imaginación de Europa se dejó dominar por las fantasías de Descartes. Como dice Quine: «Los epistemólogos soñaban en una filosofía primera, más firme que la ciencia y que serviría para justificar nuestro conocimiento del mundo exterior» [9]. Pero ¿por qué, de repente, todo el mundo comenzó a tener el mismo sueño? ¿Por qué la teoría del conocimiento llegó a ser algo más que el lánguido ejercicio acadé-

[8] «Desde el punto de vista de la filosofía medieval, Descartes desempeña el papel del *indisciplinatus* —alguien que se precia de insistir, independientemente de la disciplina en cuestión, en el mismo grado de certeza, por muy inconveniente que resulte. En una palabra, Descartes ya no reconoce un intermedio entre lo verdadero y lo falso; su filosofía es la eliminación radical de la idea de 'lo probable'». (Étienne Gilson, *Études sur le Role de la Pensée Médiévale dans la Formation du Système Cartésien* [París, 1930], página 235).

[9] W. V. O. Quine, «Grades of Theoreticity», en *Experience and Theory,* ed L. Foster y J. W. Swanson (Amherst, Mass., 1970), pág. 2.

mico de presentar una réplica a Sextus Empiricus? El sueño de una filosofía primera más firme que la ciencia es tan antiguo como *La República,* y podemos estar de acuerdo con Dewey y Freud en que por debajo de la religión y del platonismo subyacen los mismos impulsos primordiales. Pero eso no nos explica por qué alguien puede pensar que la filosofía primera consiste, ni más ni menos, que en la epistemología.

Quizá sea un poco abusivo forzar las palabras de Quine en este sentido. Sin embargo, me permito hacerlo porque creo que para entender la filosofía moderna hace falta romper con la tradición más radicalmente de lo que quiere Quine, o de lo que hace falta para sus objetivos. Su despreocupado «No arrojemos la epistemología —hagámosla psicología» es totalmente razonable si nuestro objetivo es demostrar qué se puede salvar del empirismo una vez que nos hemos deshecho de los dogmas. Pero si queremos saber por qué alguien pensó que valía la pena, y mucho menos que era apasionante o moralmente obligatorio, ser empirista, hemos de dar marcha atrás en todo este asunto e insistir en preguntas que Quine puede dejar de lado sin problemas. Para conseguir mejor este distanciamiento, examinaré ahora algunas de las cosas que Quine dice sobre la psicología. Quiero hacer ver cómo cualquier descubrimiento psicológico del tipo en que él piensa quedará muy lejos de cualquier interés por los fundamentos de la ciencia o por la relación entre teoría y evidencia.

La conexión entre psicología empírica de la percepción y epistemología empirista está, pienso yo, en el uso poco riguroso de palabras como «evidencia», «información» y «testimonio». Este uso permite a Quine decir cosas como «Las terminaciones de los nervios... son el lugar de entrada de la información no procesada sobre el mundo» [10] y «Son las estimulaciones de nuestros receptores sensoriales las que más oportunamente deben considerarse como la entrada a nuestro mecanismo cognitivo» [11]. Supongamos que hacemos estas preguntas: ¿Podría descubrir la psicología que no es en la retina (primeras células nerviosas que se ven afectadas por los rayos de luz) donde comienza a procesarse la información? ¿Podría descubrir que todo lo que se da hasta el último punto no es información sino simplemente electricidad? Probablemente no, pues es difícil ver qué debe tomarse como criterio experimental de «información» o «procesamiento». Sin embargo, Quine escribe como si *pudieran* darse es-

---

[10] *Ibíd.,* pág. 3.
[11] Quine, «Epistemology Naturalized», en *Ontological Relativity,* pág. 84.

tos criterios. Observa que la epistemología ha estado siempre dividida entre dos criterios de lo que tiene carácter de «dato»: «la proximidad causal con el estímulo físico» y «el foco de la conciencia». Pero, dice:

> El dilema se disuelve, y se alivia la tensión, cuando abandonamos el sueño de una filosofía primera más firme que la ciencia. Si estamos buscando únicamente el mecanismo causal de nuestro conocimiento del mundo externo, y no una justificación de ese conocimiento en términos anteriores a la ciencia, podemos contenrarnos en definitiva con una teoría de la visión al estilo de Berkeley basada en las manchas de color en un campo bi-dimensional... Podemos considerar al hombre como una caja negra en el mundo físico, expuesta a fuerzas estimuladoras externamente determinables en cuanto a su entrada y que produciría un testimonio externamente determinable sobre el mundo exterior, en cuanto a su entrada. No importa cuál de las acciones internas de la caja negra es la que aparece teñida de conciencia [12].

Pero si olvidamos la justificación y buscamos mecanismos causales, es seguro que no vamos a hablar de manchas de color en un campo visual bi-dimensional. No servirá de nada la distinción entre lo que es dado y lo que es fruto de la inferencia, y no habrá ninguna necesidad de la idea de un «campo visual» que sostenga a lo primero. Podemos hablar de manchas irradiadas en una retina bi-dimensional o de pulsaciones del nervio óptico, pero esto será cuestión de elegir una caja negra, no de descubrir piedras de toque de la investigación. Quine disuelve un dilema limitándose a cambiar el motivo de la investigación. Si sólo estuviéramos interesados en los mecanismos causales, nunca habríamos dado vueltas en la cabeza a la conciencia. Pero los epistemólogos que tuvieron el sueño que describe Quine no sólo tenían interés en los mecanismos causales. También les interesaba hacer, por ejemplo, una distinción injuriosa entre Galileo y los profesores que se negaban a mirar por su telescopio.

Si no hay realmente criterios experimentales para determinar dónde se dan los datos reales, la sugerencia de Quine de que abandonemos la idea de «datos sensoriales» y hablemos causalmente de terminaciones nerviosas y epistemológicamente de oraciones de observación [13] no resuelve un dilema que ha atormentado a la episte-

---

[12] Quine, «Grades of Theoreticity», en *Experience and Theory*, págs. 2-3.
[13] *Ibíd.*, pág. 3.

mología. Más bien, lo que hace es dejar que la epistemología se vaya marchitando. Si tenemos la psicofisiología para hacerse cargo de los mecanismos causales, y la sociología y la historia de la ciencia para observar las ocasiones en que se invocan o se eluden las oraciones de observación al construir o desmontar teorías, entonces la epistemología no tiene nada que ver. Podríamos pensar que este resultado está en la línea de Quine, pero de hecho va en contra de la misma. La resistencia se aprecia con toda claridad cuando el autor reprende a Polanyi, Kuhn y Hanson por querer abandonar totalmente la idea de observación [14]. Quine considera que ésta es una idea perfectamente válida, y quiere reconstruirla en términos de intersubjetividad. Define la «oración de observación» como «aquella sobre la cual todos los hablantes de la lengua dan el mismo veredicto cuando se de la misma estimulación concurrente. Dicho en forma negativa, una oración de observación es aquella que no es sensible a diferencias en la pasada experiencia dentro de la comunidad del habla» [15]. Quine piensa que, excluyendo a los ciegos, a los locos y algunos «desviados ocasionales» (pág. 88n), podemos saber cuáles son las oraciones que «dependen de la estimulación sensorial presente y no de ninguna información almacenada fuera de la que entra en la comprensión de la oración» (pág. 86). Esto equivale a definir «estimulación sensorial presente» en términos del carácter no controvertido de ciertas oraciones. Quine piensa que esto mantiene la fuerza de la idea empirista al mismo tiempo que es un abandono de las ideas sobre el significado que van asociadas a la «idea 'idea'».

Creo que Quine tiene razón al seguir esta línea para conservar lo que había de cierto en el empirismo, pues con ello deja perfectamente claro que si hay algo que «reemplace» a la epistemología es la historia y sociología de la ciencia, y no la psicología, ciertamente. Pero no es ésta la razón de Quine. Veamos otro pasaje sobre los «datos»:

> Ahora se puede establecer lo que debe contar como observación en términos de la estimulación de los receptores sensoriales, caiga donde caiga la conciencia.
>
> En los pasados tiempos anti-psicológicos se discutía la cuestión de la prioridad epistemológica. ¿A qué es anterior epistemológicamente algo? ¿Son las Gestalt anteriores a los átomos sensoriales...? Ahora que tenemos la posibilidad de recurrir a la estimulación física, el problema desaparece: A es epistemológicamente anterior a B si A está casualmente más próximo

[14] *Ibíd.*, pág. 5; cfr. «Epistemology Naturalized», pág. 87.
[15] Quine, «Epistemology Naturalized», págs. 86-87.

que B a los receptores sensoriales. O, lo que en algunos sentidos resulta todavía mejor, basta con hablar explícitamente en términos de proximidad causal a los receptores sensoriales y dejar de hablar de prioridad epistemológica (págs. 84-85).

Lo que resulta sorprendente es que hayamos definido la «oración de observación» en términos de *consensus gentium;* podemos distinguir la observación de la teoría sin saber o tener interés en qué partes de nuestro cuerpo son los receptores sensoriales, y mucho menos en qué punto de los nervios comienza el «procesamiento». No necesitamos ninguna explicación psicofisiológica de los mecanismos causales para aislar lo que se puede admitir de forma intersubjetiva —hacemos esto mismo en la conversación normal. Por eso es de suponer que la psicología no tiene nada que decirnos sobre la proximidad causal que valga la pena saber a quienes desean continuar con la «epistemología en un marco psicológico». Dicho en otras palabras, una vez que hemos descubierto las oraciones de observación por procedimientos conversacionales más que neurológicos, el continuar la investigación de «cómo se relaciona la evidencia con la teoría» parecería ser cuestión de Polanyi, Kuhn y Hanson. ¿Qué podría añadir la psicología a sus explicaciones de cómo forman y rechazan las teorías los científicos? Quine dice que

> ... algunos filósofos de la ciencia iconoclastas se han dedicado a poner en duda la idea de observación justamente ahora, cuando ya ha dejado de constituir un problema. Creo que se trata de una reacción retrasada contra el carácter dudoso de la antigua idea de dato. Ahora que hemos descartado el antiguo sueño de una filosofía primera, debemos más bien alegrarnos por nuestro nuevo acceso a conceptos no problemáticos. Uno de ellos es el de entrada nerviosa, y otro el de oración de observación tal como acabamos de describirla [16].

Pero estos conceptos, suponiendo que sean no-problemáticos, no son nuevos. Como la electricidad, la entrada nerviosa no es nueva; como la «información», es problemática. La idea de «oración de observación» tal como la define Quine es tan antigua como el primer abogado que preguntó a un testigo: «¿Qué es lo que vio realmente?» Si de algo hemos de alegrarnos es de que ya no nos hacemos ciertas preguntas —no de que hayamos encontrado algo nuevo que hacer o algunos nuevos términos en qué pensar. Quine nos ha dicho que dejamos de preocuparnos por la conciencia cuando

[16] Quine, «Grades of Thoreticity», págs. 4-5.

abandonamos la reconstrucción racional. Pero parece que él la vuelve a traer al explicar la observacionalidad en términos de intersubjetividad. Por eso, debería o dejar que Polanyi, Kuhn y Hanson digan que la «observación» consiste simplemente en aquello en que podemos ponernos de acuerdo en la actualidad, o demostrar cómo los descubrimientos psicológicos pueden dar algo más de importancia a esta idea. Si no puede, entonces defender «la dependencia de la estimulación sensorial presente» en términos de intersubjetividad será simplemente invocar un tratamiento de honor epistemológico para nada.

Al hablar de Quine he insistido en sus propias palabras con una literalidad algo fatigosa. Es probable que a Quine no le preocupe el destino de la palabra «epistemología». Lo que quizá le preocupe es su posición deweyana según la cual la ciencia y la filosofía tienen carácter continuo —lo que no debe entenderse como que tenga métodos o materias diferentes. No quiere renunciar a la forma de hablar oxoniana de la «filosofía como análisis de conceptos», y asocia a Wittgenstein y al «positivismo terapéutico» con esta forma de hablar. Ya he dicho que insistimos más en lo que Dewey y Wittgenstein tienen en común —su opinión de que los filósofos modernos han realizado al mismo tiempo una búsqueda natural de la comprensión y una búsqueda no natural de la certeza. Según esta concepción, están igualmente descaminados los temores y esperanzas que la psicología ha suscitado en diversos momentos entre los filósofos [17]. Decir con Wisdon y Bouwsma que la «epistemología» es una reunión de preocupaciones obsesivas por la certeza que debe eliminarse mediante la terapia, y decir con Quine que los impulsos epistemológicos deben satisfacerse con los resultados psicológicos, pueden considerarse como formas de decir: Podemos tener psicología o no tener nada.

Si no fuera porque una de mis preocupaciones en este libro es

---

[17] Véase Ludwig Wittgenstein, *Philosophical Investigations* (Londres, 1953), página 232, sobre «la confusión y esterilidad de la psicología»: «La existencia del método experimental nos hace pensar que tenemos los medios para solucionar los problemas que nos inquietan; sin embargo, problema y método se hacen caso omiso». Véase también las advertencias de Dewey sobre el movimiento que llegaría a convertirse en la «psicología conductista» que suscita la admiración de Quine: «El dualismo anterior entre sensación e idea se repite en el actual dualismo de las estructuras y funciones periféricas y centrales; el anterior dualismo de cuerpo y alma encuentra un eco claro en el actual dualismo del estímulo y respuesta...»; y en otro lugar: «... la sensación como estímulo no significa ninguna *existencia* física concreta. Significa simplemente una función, y hará que su valor cambie según la tarea especial que se deba realizar» (The Reflex Arc Concept in Psychology», en *The Early Works of John Dewey*. vol. 5 [Carbondale, 1972], págs. 96, 107).

preguntar por qué tenemos en nuestra cultura un fenómeno como el de la «filosofía», podríamos dejar las cosas en este punto. Pero para esta cuestión histórica la diferencia entre nada y psicología es importante. Dewey subrayó los motivos religiosos y sociales que había tras los aspectos no-«científicos» de la filosofía, y unía su insistencia sobre la continuidad entre filosofía y ciencia con una distinción ofensiva entre lo que había sido la filosofía y lo que debería llegar a ser. Quine se resiste a juguetear con una falacia genética, y está alegremente dispuesto a considerarse a sí mismo y a Locke como compañeros de investigación de «la relación entre teoría y evidencia»: Locke, piensa él, se dejó llevar por el mal camino por culpa de una mala teoría del significado, pero los modernos podemos volver al buen camino (hacia la psicología) gracias a una teoría adecuada. Pero este optimismo oculta precisamente lo que es más importante para la comprensión histórica: La preocupación de Locke por la sugerencia escéptica de que nuestros modos subjetivos de aprehensión pueden ocultarnos la realidad, y la negativa de Quine a dejarse molestar lo más mínimo por el escepticismo.

El alejamiento de Quine de las preocupaciones escépticas se manifiesta en su asimilación de los elementos de la experiencia con los elementos del conocimiento, y de la explicación con la justificación. La psicología, al encontrar los elementos de la experiencia, explica el conocimiento. La epistemología, encontrando (supuestamente) los elementos del conocimiento, justifica el conocimiento no elemental. Nadie desaría justificar el «conocimiento humano» (en oposición a una teoría o informe concretos) a no ser que estuviera asustado por el escepticismo. Nadie asimilaría la epistemología a la psicología a no ser que tuviera tan poco temor al escepticismo como para considerar que el intento de «buscar una base al conocimiento humano» no pasaba de ser una broma sin importancia. Por eso, aunque podamos estar plenamente de acuerdo con Quine en que si se van a hacer descubrimientos sobre el conocimiento humano es probable que procedan de la psicología, podemos también simpatizar con la opinión que Quine atribuye a Wittgenstein: Lo que hay que hacer con la epistemología es «curar a los filósofos de la alucinación de que hay problemas epistemológicos». Esta terapia no separa la filosofía de la ciencia: supone que la filosofía es sólo sentido común o ciencia que se moviliza para proporcionar «advertencias con un objetivo concreto» [18].

[18] Wittgenstein, *Philosophical Investigations,* pág. 50.

3. Los estados psicológicos en cuanto explicaciones auténticas

Para conseguir una teoría psicológica que dijera algo sobre la relación entre teoría y evidencia necesitamos, cuando menos, una teoría que reproduzca «internamente» la habitual justificación «pública» de las afirmaciones mediante circunstancias y otras afirmaciones. Necesitamos, en otras palabras, realidades mentales que puedan tener con las afirmaciones públicas y entre sí las mismas relaciones que tienen las premisas y conclusiones en el habla, que los testimonios de los testigos de la acusación en los tribunales, etc. Pero siempre que se propone una teoría psicológica que responda a esta necesidad, es probable que se oiga la voz de «regreso infinito». Así, Malcolm dice:

> Si decimos que una persona sabe que algo que tiene delante de él es un perro viendo que la criatura «encaja» en su Idea de lo que es un perro, entonces debemos preguntar: «¿Cómo sabe que esto constituye un ejemplo de encajar?» ¿Qué es lo que orienta su juicio en esto? ¿No necesita una Idea de segundo orden que le haga ver qué es eso de encajar con una Idea? Es decir, ¿no necesitará un modelo de *encajar?*... Se ha producido un regreso hacia el infinito y no se ha explicado nada [19].

La dificultad, dada a conocer por Ryle, es que si no nos conformamos con considerar «Él lo ve» como justificación suficiente del conocimiento por parte del hombre de que hay un perro ante él, no podremos aceptar como justificación ninguna otra cosa. En la medida en que una explicación mentalista proporciona meramente una explicación causal del reconocimiento por la vista, parece que no responde a la pregunta «¿Cómo lo sabe?». No nos dice nada sobre la *evidencia* de este hombre en favor de su opinión, sino únicamente de cómo ha llegado a adquirirla. Por otra parte, en la

[19] Malcolm, «Myth of Cognitive Processes», pág. 391. Compárense Ryle, *The Concept of Mind* (Nueva York, 1949), capítulo 7, y Wittgenstein, *Philosophical Investigations,* págs. 213-215. Véase también John Passmore, *Philosophical Reasoning* (Londres, 1961), capítulo 2 («The Infinite Regress Argument»), donde examina el uso que hace Ryle del argumento. Yo he comparado el empleo anticartesiano que Wittgenstein y Peirce hacen de este argumento, en «Pragmatism, Categories, and Language», *Philosophical Review,* 49 (1961), 197-223.

medida en que ofrece una justificación de la pretensión de conocimiento público y original, constituye una ocasión para seguir buscando una nueva justificación.

Fodor critica la afirmación de Ryle de que no hay nada «paramecánico» que pueda mejorar nuestra comprensión del reconocimiento perceptivo, y comenta que «la atrayente simplicidad de la postura de Ryle se consigue haciendo una petición de principio y dando por solucionadas precisamente aquellas preguntas que las teorías de la percepción y el aprendizaje han intentado responder tradicionalmente»[20]. Luego trata de demostrar que es inútil intentar responder a estas preguntas con «alguna historia simple sobre las asociaciones aprendidas»:

> Pero si lo que tienen en común las distintas formas de interpretar «Lillibulero» es algo abstracto, parece desprenderse que el sistema de expectaciones que constituye la receta propia para oír la canción ha de ser abstracto en ese mismo sentido...
>
> ...Las expectaciones relevantes han de ser complicadas y abstractas, pues las identidades perceptivas son sorprendentemente independientes de las uniformidades físicas entre los estímulos. Como es precisamente esta «constancia» perceptual lo que los psicólogos y epistemólogos han supuesto tradicionalmente que necesitará inferencias inconscientes y otras transacciones para-mecánicas para su explicación, parece relevante hacer notar que el tratamiento de Ryle da por respondidas todas las cuestiones que plantea la constancia (págs. 377-378).

Podemos estar de acuerdo con Fodor en que si *hay* «cuestiones que plantea la constancia», Ryle ha incurrido en una petición de principio. Pero Ryle podría replicar fácilmente que es la idea de «expectaciones complicadas y abstractas» (por ejemplo, un conjunto de inferencias inconscientes que implican hacer referencia a unas reglas o a unos paradigmas abstractos) lo que hace que se planteen cuestiones. Quizá lo único que nos haga preguntar «¿Cómo se hace?» sea la imagen de un hombrecillo dentro de la mente, que aplica reglas formuladas en términos no verbales pero a pesar de ello «abstractos». Si no nos hubieran impuesto esta imagen, podría decir Ryle, responderíamos con algo parecido a «Sólo es posible por el hecho de tener un sistema nervioso complicado —indudablemente, algún día un fisiólogo nos dirá cómo funciona». En otras palabras,

---

[20] Fodor, «Could There Be a Theory of Perception?», *Journal of Philosophy*, 63 (1966), 375.

la idea de «modelos» no fisiológicos no surgiría si no tuviéramos ya a mano todo el cúmulo de trucos cartesianos.

Es posible reformular esta respuesta con algo más de precisión. Supongamos que estamos de acuerdo con Fodor en que el reconocimiento de la semejanza entre diferencias potencialmnte infinitas es el reconocimiento de algo «abstracto» —la Lillibulereidad, por ejemplo. ¿Qué significa que «la receta propia para oír la canción ha de ser abstracta en ese mismo sentido»? Probablemente, que ha de ser capaz de distinguir la semejanza entre diferencias potencialmente infinitas. Pero entonces no sirve de nada la idea de «receta *no*-abstracta», pues *cualquier* receta debe ser capaz de realizar esto. También son potencialmente infinitas las posibles variaciones cualitativas entre los ingredientes de un lote de galletas de chocolate. Así pues, si queremos hablar de «conjuntos complicados de expectaciones» (o «programas» o «sistemas de reglas»), estaremos hablando siempre de algo «abstracto» —igual de abstracto, de hecho, que la característica cuyo reconocimiento (o la tarea cuya ejecución) deseamos explicar. Pero entonces nos encontramos ante un dilema: o bien la adquisición de estos conjuntos de expectaciones o reglas requiere la postulación de nuevos conjuntos de expectación o reglas, o no son adquiridas. Si optamos por la primera opción, el regreso hacia el infinito de Malcolm será producido en este caso por el principio de Fodor de que el conocimiento de lo abstracto requiere el uso de lo abstracto, pues lo que es válido respecto al reconocimiento debe serlo también para la adquisición. Si nos inclinamos por la segunda posibilidad, parece que volvemos a Ryle: decir que las personas tienen una capacidad no adquirida de reconocer la semejanza entre diferencias infinitas difícilmente puede considerarse como algo explicativo sobre «las cuestiones que plantea la constancia».

Por eso, puede concluir Ryle, estas cuestiones o son de carácter «conceptual» y tratan de las condiciones suficientes para la aplicación normal de términos como «reconocer», o se refieren a los mecanismos fisiológicos. Este último tipo de cuestión no implica problemas sobre retrocesos, pues nadie piensa que la «constancia» exija postular mecanismos «abstractos» en las células fotoeléctricas o diapasones. Sin embargo, ¿hay alguna diferencia entre el do medio y la «Lillibulereidad», salvo que al primero lo hemos apodado «cualidad acústica concreta» y a la segunda «semejanza abstracta»? Podríamos especificar mil rasgos accidentales (timbre, volumen, presencia de luz, color del objeto que emite el sonido) que ignora el diapasón, lo mismo que ocurre con el reconocedor-de-Lillibulero.

Como la distinción abstracto-concreto es tan relativa a una determinada base de datos como la distinción complejo-simple, parece como si, al decir que la explicación psicológica requiere una referencia a entidades abstractas, estemos afirmando simplemente que para explicar el tipo de cosas que hacen los mamíferos hace falta referirse a tipos de cosas diferentes —*categóricamente* diferentes— de explicar qué pueden hacer las amebas, los diapasones, los átomos de cesio y las estrellas. Pero, ¿cómo sabemos eso? Y ¿qué significa aquí «categóricamente»? Una vez más, Ryle puede decir que si no tuviéramos ya la imagen cartesiana (de un Ojo Interior que contempla las reglas escritas en los muros del escenario mental), no sabríamos qué hacer con la afirmación.

Y basta ya por lo que se refiere a la fuerza del argumento del retroceso infinito. Consideremos ahora el tipo de réplica al mismo que haría alguien como Dodwell, que afirma que la construcción de modelos de carácter no-fisiológico no es, a priori, ni buena ni mala, y que deja justificarse por sus frutos. Dodwell está impresionado por la analogía entre cerebros y computadoras: «La mayor influencia sobre las ideas de los psicólogos sobre los procesos cognitivos es en la actualidad el nexo de conceptos que se ha elaborado para la programación por computadora» [21]. Sin embargo, admite lo siguiente:

> Podría argumentarse que la analogía con la computadora es superficial, pues un programa se limita a codificar un conjunto de operaciones que son parecidas a las operaciones cognitivas, pero no *explica* el pensamiento más de lo que pueda hacerlo escribir un conjunto de reglas para resolver problemas aritméticos... Decir que un programa de computadora puede «explicar» el pensamiento tendría aproximadamente la misma fuerza que decir que un conjunto de fórmulas lógicas «explica» las leyes de la argumentación deductiva correcta (págs. 371-372).

A este argumento responde que la analogía computacional sólo tiene fuerza una vez que se distinguen niveles:

> ... las explicaciones de lo que ocurre en la resolución de problemas por una computadora pueden hacerse en distintos niveles... La *aplicación* de un programa debe explicarse en términos del hardware de la computadora, lo mismo que la aplicación del pensamiento ha de explicarse, en cierto sentido, por procesos que ocurren realmente en el sistema nervioso central.

[21] Dodwell, «Is a Theory of Conceptual Development Necessary?», página 370.

> Las subrutinas por las que se hacen las computadoras particulares se pueden explicar haciendo referencia al «lenguaje de máquina» y a los algoritmos paso-por-paso mediante los cuales se encuentran las soluciones... El principio de la operación de la subrutina no se debe entender y explicar limitándose a examinar el hardware, de la misma manera que no se podrían entender las tablas de multiplicar examinando el cerebro. De forma semejante, la comprensión de cómo funcionan las mismas subrutinas no explica el principio de resolución de problemas en términos de una secuencia de pasos... Para eso, hay que considerar el proceso ejecutivo, que en la máquina encarna la organización general y el objetivo del programa, y en el ser humano una «orientación hacia un objetivo» que no se entiende con tanta claridad (pág. 372).

La importancia de los niveles queda ilustrada, por ejemplo, por el hecho de que la experimentación puede darnos razones para decir que reconocemos las formas visuales por un proceso de emparejamiento de plantillas más que por un proceso de extracción de rasgos (pág. 379). Decir esto no es hacer una observación ni «conceptual» (sobre «el proceso ejecutivo») ni «fisiológica» (sobre el «hardware»), pero, sin embargo, puede ser verdaderamente explicativa. La idea de «subrutina» parece ofrecernos precisamente lo que necesita la psicología —una explicación de para qué podría servir el terreno intermedio entre el sentido común y la fisiología.

Pero, ¿cómo nos ayuda esta idea en el argumento del retroceso infinito? Probablemente, Malcolm y Ryle insistirían en que las «plantillas» o ideas abstractas de los rasgos extraídos (según el modelo que se elija) producen los mismos problemas que la «constancia» que se supone deben explicar. Pero Dodwell puede responder que sólo lo harían si se suponía que eran respuestas a preguntas generales del tipo «¿Cómo es posible la abstracción (reconocimiento, constancia)?». Para estas preguntas, puede decir él, no hay ***ninguna*** respuesta, como no sea la observación sin sentido de que la naturaleza ha elaborado el hardware adecuado para que se realice la tarea. Efectivamente, cualquiera de los modelos de Dodwell será verdaderamente antropomórfico, en el sentido de pensar en un pequeño inferidor dentro del cerebro que comprueba sus plantillas o marca con una señal los rasgos. La capacidad de abstracción o reconocimento de este inferidor serán tan problemáticas como la de su anfitrión, y no lo serán menos si decimos que es una maquinita en vez de un hombrecillo [22]. En este punto los modelos antropomórficos del pro-

[22] Algunos psicólogos pondrían esto en duda. Gregory, citando en sentido

gramador cuando dice «la máquina no entenderá el problema si utiliza los signos polacos, pues lo único que sabe...» Quejarse de que las «plantillas» —como las «ideas» de Locke— son una reduplicación del 'explanandum' es como afirmar que las partículas que componen el átomo de Bohr son una reduplicación de las bolas de billar cuya conducta ayudan a explicar. Resulta provechoso postular pequeñas bolas de billar dentro de las grandes, así que ¿por qué no postular personas pequeñas dentro de las grandes (o ratas pequeñas dentro de las ratas grandes)? Cada uno de estos «modelos» va acompañado, como dice Sellars, de un «comentario» que enumera los rasgos de la realidad modelada «abstraídos de» en el modelo[28]. Parece razonable señalar que el comentario implícito en todos los modelos antropomórficos de la psicología va más o menos por estas líneas:

> Mientras nos mantengamos en el nivel de las subrutinas, tendremos libertad para hablar antropomórficamente sobre las inferencias y otras operaciones realizadas «inconscientemente» por la persona, o realizadas (ni «conscientemente» ni «inconsciente-

aprobativo la idea de «inferencias inconscientes», de Helmholtz, implicadas en la percepción, dice que

> debemos tener claro que no hay ningún «hombrecillo interior» que haga el argumento, pues esto lleva a dificultades filosóficas inadmisibles. Helmholtz no pensaba en esto, ciertamente, pero su expresión «inferencias inconscientes» y su descripción de las percepciones como «conclusiones inconscientes» quizá hizo pensar, entonces, a personas no familiarizadas con los computadores, en algo tan inaceptable. Pero nuestro conocimiento de los computadores debería eliminar toda tentación de confusiones de este tipo. Ya no pensamos que la inferencia sea una actividad exclusivamente humana que implique conciencia. (*The Intelligent Eye* [Nueva York, 1970], página 30).

Considero un error decir que el hombrecillo lleve a «dificultades filosóficas inadmisibles», pues no creo que las maquinitas sean menos «conscientes» que los hombrecillos. Adoptar lo que Dennett llama «actitud intencional» hacia las series de transistores o neuronas es hablar de ellos como hablamos de los seres conscientes, y el añadir «pero, por supuesto, no son *realmente* conscientes» parece que es decir meramente que no tenemos ninguna responsabilidad moral hacia ellos. No podemos ni investigar cuáles de estas series están, en frase de Quine, «impregnadas de conciencia» ni *descubrir* que puedan hacer inferencias seres que no estén impregnados en esa forma. La familiaridad con los computadores no produce tal descubrimiento; sólo hace más corriente y casual la atribución de la postura intencional.

[28] Véase Wilfrid Sellars, *Science, Perception and Reality* (Londres y Nueva York, 1963), pág. 182, donde habla de los «comentarios» y las págs. 192 y ss., donde trata de las impresiones sensoriales rojas.

> mente») por los centros del cerebro u otros órganos de los que se habla como si fueran personas. El uso de estas expresiones no nos obliga a atribuir a los centros cerebrales intelecto ni carácter, lo mismo que el hablar de la «impresión sensorial roja» como factor común en varias ilusiones no nos obliga a admitir la existencia de algo que sea a la vez «interior» y rojo. Pero una vez que abandonamos el nivel de la «subrutina» y pasamos al nivel del hardware, el antropomorfismo está fuera de lugar.

Para ver la fuerza de este comentario, supongamos que hubiera una clase especial de corriente nerviosa que recorriera el nervio óptico cuando, y sólo cuando, la teoría psicológica previera la ocurrencia de una impresión sensorial roja (y lo mismo en todas las demás situaciones perceptivas). Si supiéramos este hecho, dejaríamos sin más la explicación de la «subrutina» para pasar directamente al hardware. La idea de «impresión sensorial» ya no tendría un papel que desempeñar (a no ser que hubiera otras entidades teóricas postuladas por la teoría psicológica y que requirieran esta noción para su explicación). Si las cosas resultaran así de sencillas, entonces la analogía de la «computadora» ya no parecería especialmente relevante —lo mismo que no lo parece en el caso de los animales unicelulares, donde el paso de la conducta a la fisiología es demasiado pequeño para que parezca tener sentido la idea de «niveles».

Esto equivale a decir que si la fisiología fuera más simple y más obvia de lo que es, nadie habría sentido la necesidad de la psicología. Esta conclusión puede parecer extraña, especialmente a la luz de la observación de Dodwell (citada más arriba) de que «el principio de la operación de la subrutina no se debe entender y explicar limitándose a examinar el hardware, de la misma manera que no se podrían entender las tablas de multiplicar examinando el cerebro»[24]. Pero esta observación es gravemente engañosa. Representa una confusión entre lo evidente:

> Si no supiéramos lo que era la multiplicación, el mirar al cerebro no nos lo diría nunca

y lo dudoso :

---

[24] Fodor ha apuntado también que la distinción en psicología entre análisis «funcional» (o del «programa») y análisis «mecánico» (o del «hardware») es irreductible y no sólo cuestión de conveniencia. Véase sus «Explanations in Psychology», en *Philosophy in America*, ed. M. Black (Ithaca, 1965), página 177. He escrito contra esta idea en «Functionalism, Machines, and Incorrigibility», *Journal of Philosophy*, 69 (1972), 203-220.

> Si supiéramos qué era la multiplicación, no podríamos saber que alguien estaba haciendo una determinada operación de multiplicar con sólo mirar su cerebro.

Esto último es dudoso porque no sabemos precisamente si hay o no hay parámetros neurofisiológicos muy simples asociados con ciertas operaciones mentales. Es sumamente improbable que los haya, pero no hay ninguna razón a priori para que una adecuada exploración-del-cerebro-cum-microscopio no haga ver algo de que el observador adiestrado pueda informar con estas palabras: «Ah, estás multiplicando cuarenta y siete por veinticinco» (y acertar en todas las ocasiones). Más en general, la cuestión de qué se explica mejor en términos de hardware y qué en términos de programa depende totalmente de lo adecuado que resulte ser el hardware y de cómo esté diseñado. La adecuación y la calidad del diseño se refieren a la elección de vocabulario y nivel de abstracción —pero también hace eso la misma distinción hardware-software [25]. Dada la clase adecuada de hardware y los parámetros adecuados, *es* ciertamente posible «entender y explicar el principio de la subrutina con sólo examinar el hardware». En realidad, podemos imaginar máquinas en las que sería más fácil averiguar qué es capaz de hacer la máquina abriéndola y mirando que leyendo el programa.

Como el cerebro *no* es, casi con toda seguridad, esta máquina, estamos ante una cuestión de principio, pero un principio que tiene importancia filosófica. Demuestra que la distinción entre psicología no es una distinción entre dos objetos materiales distintos en un sentido más fuerte que pueda serlo, por ejemplo, la distinción entre química y física. Podría haber resultado que los fenómenos químicos del tipo de la formación de compuestos no hubiera tenido nada que ver con la estructura submicroscópica de los elementos en cuestión. Pero de hecho tienen que ver, y por eso ahora el que utilicemos términos físicos o químicos para explicar una reacción es cuestión de comodidad o pedagogía. Si resulta que la fisiología tiene tanto que ver con el multiplicador como los electrones con las explosiones, la distinción psicología-fisiología será igualmente pragmática. Por eso se puede reafirmar la paradójica conclusión presentada anteriormente —que si la fisiología hubiera sido más obvia no habría

---

[25] Sobre esta clase de relatividad, véase William Kalke, «What Is Wrong with Fodor and Putnam's Functionalism», *Nous,* 3 (1969), 83-94. Véase la crítica al artículo de Kalke y al mío paralelo (citado en la nota 24) en B. J. Nelson, «Functionalism and the Identity Theory», *Journal of Philosophy,* 73 (1976), 379 y ss.

aparecido la psicología. En realidad, podemos reforzarla y decir que si hubiera sido más fácil entender el cuerpo, nadie habría pensado que teníamos una mente[26].

Es ya hora de resumir esta forma de hacer frente al argumento del retroceso infinito. El punto central es precisamente que las entidades explicativas postuladas por los psicólogos reduplican los problemas de los 'explananda' únicamente cuando estos problemas son de por sí malos —por ejemplo, «¿Cómo es posible el reconocimiento?». Filósofos como Malcolm y Ryle están acostumbrados a malas respuestas filosóficas a malas preguntas filosóficas: «¿Cómo es posible el movimiento? —como actualización del potencial en cuanto potencial»; «¿Por qué sigue leyes la naturaleza? —por la benevolencia y omnipotencia de Dios». Por consiguiente, suelen ver cuestiones de este tipo ocultas incluso detrás de programas de investigación muy específicos y limitados. No siempre se equivocan, pues los psicólogos ofrecen todavía en algunas ocasiones su último «modelo» como solución de antiguos problemas filosóficos[27]. Pero su-

---

[26] Como espero haber aclarado en el capítulo segundo, no quiero decir con esto que no debiéramos haber pensado en nosotros mismos en cuanto que tenemos creencias y deseos, vemos, inferimos, etc. Sino que no deberíamos aceptar las ideas de «un entendimiento activo separable», una «sustancia inmaterial» cartesiana» o las ideas lockeanas. Nuestro concepto de mente habría estado mucho más cerca del de Ryle o Aristóteles que del concepto cartesiano que tenemos actualmente.

[27] Véase, por ejemplo, la (Introducción» de Seymour Papert a Warren S. McCulloch, *Embodiments of Mind* (Cambridge, Mass., 1965). Papert, al explicar la importancia de la obra de McCulloch, nos dice que «ya no debemos sentirnos atrapados por el dilema» de

> una escisión... entre la psicología, que estaba basada en el mecanismo pero era incapaz de llegar a las propiedades complejas del pensamiento, y la filosofía, que se tomaba muy en serio las propiedades del pensamiento pero no se contentaba con ningún mecanismo imaginable (pág. xiv).

La idea que va a resolver este dilema tiene como «principal paso conceptual»

> el reconocimiento de que una gran cantidad de situaciones diferentes que implican la regulación teleonómica de la conducta en los sistemas mecánicos, eléctricos e incluso sociales deberían entenderse como manifestaciones de un solo fenómeno básico: la vuelta de información para formar un bucle central cerrado (pág. xvi).

Papert supone aquí que las «propiedades del pensamiento» que preocupaban a los filósofos eran las relevantes a la posesión de un propósito. Pero ni la distinción entre justificación y explicación causal ni la que hay entre conciencia y la carencia de la misma —que son las dos grandes formas de distinguir la

pongamos que los modelos como los que tiene presentes Dodwell —propuestas sobre subrutinas que no pueden someterse a introspección (como el «proceso ejecutivo») ni son fisiológicamente descifrables (como el «hardware»)— no se consideran ni como aportaciones para resolver los seudo-problemas ni como descubrimientos sobre algún tipo de entidad no físico. Entonces el argumento del retroceso infinito carece de fuerza. Tanto si aparecen como si no aparecen los correlativos de hardware para estas subrutinas, el éxito en la predicción y control de la conducta, hecho posible por el descubrimiento experimental de tales subrutinas, sería suficiente para demostrar la realidad de los objetos de la investigación psicológica [28]. La sugerencia de Dodwell es que 'nada tiene tanto éxito como el éxito' cuando se trata de establecer el carácter no mítico y «científico» de la propia materia de estudio, y esto podría ser perfectamente la última palabra al respecto.

Aplicando este punto al dilema de Ryle sobre las capacidades adquiridas y no adquiridas que he esbozado anteriormente, podemos admitir alegremente que semejante construcción de modelos debe reconocer que la naturaleza ha conectado algunas capacidades no adquiridas para que realicen operaciones mentales de orden superior. Al menos algunos de esos hombrecillos que realizan subrutinas en distintos centros cerebrales habrán tenido que estar allí desde el momento del nacimiento. Pero, ¿por qué no? Si se abandona la idea de que la psicología empírica va a hacer lo que no consiguieron realizar los empiristas ingleses —demostrar cómo una *tabula rasa* se transforma en un complicado dispositivo de procesamiento de la información gracias a los impactos sobre los órganos sensoriales periféricos— no nos sorprenderemos de que la mitad de las subrutinas del adulto estén conectadas en el cerebro del niño siguiendo instrucciones de los cromosomas. Además, no nos parecerá importante para la comprensión de la naturaleza del hombre o de su mente descubrir

---

epistemología de la psicología— se clarifica clarificando la posesión de un propósito. Y tampoco se sentirían aliviados los bergsonianos, a quienes preocupa este tema, con la contemplación de cambiadiscos automáticos.

[28] Es tentador pensar en las «variables intervinientes» postuladas por los psicólogos (con «subrutinas» escritas en términos de las mismas) como *meros* sustitutos de los procesos neurológicos no descubiertos. Lo que hacemos generalmente es suponer que cuando la neurofisiología llegue a cierto punto servirá como piedra de toque para elegir entre «modelos de la mente» psicológicos enfrentados. Pero es importante ver que aun cuando llegáramos a descubrir de alguna manera que la neurofisiología nunca llegará hasta donde habíamos esperado, esta desilusión no haría más dudosa la tarea de los psicólogos, ni «metodológicamente» ni «metafísicamente».

cuáles son precisamente las que estaban conectadas desde entonces y las que vinieron más adelante [29]. Finalmente, no nos parecerá extraño que algo «abstracto» (como una capacidad de reconocer la semejanza en la diferencia) tenga tan poco de adquirido como capacidades tan «concretas» como dar una respuesta diferencial a un do agudo. Podemos recordar simplemente que esto último es a su vez tan «abstracto» como pudiera serlo una capacidad, y no más abstracto de lo que debe ser cualquier capacidad. Toda la idea de capacidades concretas-frente-a-abstractas, que Fodor y Kant aceptan sin crítica, es semejante a la idea de lo «irreductiblemente físico» frente a lo «irreductiblemente psíquico». Nadie puede decir cómo se deben trazar estas líneas, excepto en relación con los objetivos temporales de la investigación. Pero el intento cartesiano de trazar dichas líneas de una vez por todas y los intentos «empirista» y «conductista» por «reducir» lo uno a lo otro han hecho aparecer la idea de que la investigación psicológica puede llegar a penetrar en algunos de los profundos misterios que habían confundido a los filósofos. El uso negligente del argumento del retroceso infinito por parte de Malcolm y Ryle debería entenderse, creo yo, como una reacción comprensible frente a esta idea de que la psicología puede conseguir resolver los problemas que han planteado los filósofos.

Ahora puedo ya unir el resultado de examinar el argumento del «retroceso infinito» con el de la sección anterior diciendo que la idea de los estados psicológicos en cuanto representaciones internas no plantea ninguna objección pero es muy poco interesante. Decir que los estados psicológicos son estados postulados para explicar la conducta, que todavía no sabemos cómo identificar con los estados

[29] La idea de que es importante descubrir qué es «innato» aparece en preguntas como «¿Atraviesa todo el conocimiento (el término contemporáneo es información) los órganos sensoriales o hay algún conocimiento aportado por la mente misma?» (J. J. y E. J. Gibson, «Perceptual Learning: Differentiation or Enrichment?», *Psychological Review,* 62 [1955], 32). Gibson y Gibson se toman esta pregunta kantiana con toda seriedad, e insisten en que, con perdón de Hume y Helmholtz, el aprendizaje perceptivo no es una inferencia inconsciente a partir de las huellas de la memoria, sino simplemente «mayor sensibilidad a las variables de la formación estimular» (pág. 40). Sin embargo, es muy difícil imaginar cómo un experimento podría ayudar a decidir entre esta concepción y, por ejemplo, la interpretación neo-helmholtziana, de Gregory, de los experimentos clásicos de aprendizaje perceptivo. Cfr. R. L. Gregory, *Eye and Brain* (Nueva York y Toronto, 1966), especialmente pasajes como el de la página 11: «Los sentidos no nos ofrecen directamente una imagen del mundo; más bien nos presentan pruebas para comprobar las hipótesis sobre lo que hay delante de nosotros.» Véase la exposición de Gibson hecha por Fodor, que cito y discuto brevemente en la sección 4.

fisiológicos, no es descubrir la naturaleza verdadera de la mente; no es más que volver a insistir en que no hay ninguna «naturaleza» que conocer. La analogía entre mentes y computadoras expuesta por Dodwell y Fodor es mejor que la analogía de Platón entre mente y pajareras sencillamente porque la primera evita la imagen *epistemológicamente* (no metafísicamente) engañosa de la introspección en cuanto observación de nuestro interior. La iluminación que producen estas analogías consiste en responder a lo que preocupa a los filósofos, y lo que les preocupa es la ciencia unificada, por una parte, y la «subjetividad», por la otra. Si admitimos que no se habrían producido las explicaciones mentalistas si nuestro hardware hubiera sido más claro, esto será suficiente para hacer de la distinción mente-cuerpo algo pragmático más que ontológico. A su vez, esto es suficiente para reconciliarnos con el hecho de que quizá *nunca* consigamos una explicación neurofisiológica de lo que ocurre dentro de nosotros que esté tan claramente relacionada con los estados psicológicos como la explicación del ingeniero de cómo el hardware «realiza» el programa de la computadora. Una vez que, por razones quineanas, dejamos de pensar que la posibilidad o imposibilidad de dicha explicación se puede determinar por «análisis filosófico», podemos ver que la unidad de la ciencia sólo la ponen en peligro los espíritus, no lo desconocido o lo irreductible. Lo mismo que la sugerencia de que los átomos de Demócrito y las luces brillantes de Newton no son más que sacudidas geometrodinámicas no molesta a los instintos «fisicistas» de cualquiera, tampoco lo hará la afirmación de Putnam de que nunca dispondremos de una explicación en micro-partículas de por qué las clavijas cuadradas no encajan en los agujeros redondos. La necesidad permanente de hablar en sentido mentalista sólo parece peligrosa a los filósofos que piensan que «lo mental» implica hacer referencia a los espíritus. Sellars se ocupa de ellos cuando trata el carácter dado de lo mental. Sellars demuestra que cuando hacemos introspección, para el observador no físico no hay presente ningún elemento no físico. De esta manera evita la pérdida de «objetividad científica» con que nos habían amenazado. Una vez más, las semillas de los problemas metafísicos aparecen en las dificultades epistemológicas, y específicamente en la idea de que para entender cómo tenemos derecho a estar seguros de que estamos nostálgicos tenemos que establecer una divisoria ontológica entre nostalgia y neuronas.

4. Los estados psicológicos en cuanto representaciones

Sin embargo, debemos ponernos en guardia ante los intentos de erigir una nueva barrera mental-física en la línea que separa lo abstracto de lo concreto, las representaciones de las no-representaciones, en vez de siguiendo las líneas de mayor y menor certeza. Para ver más o menos en qué consiste este intento, veamos la rehabilitación que hace Fodor de las tradicionales explicaciones empiristas de la percepción:

> El saber si los procesos psicológicos son procesos computacionales es, supongo yo, una cuestión empírica. Pero si lo son, lo que tiene que ocurrir en la percepción es que una descripción del entorno que *no* esté formulada en un vocabulario cuyos términos designen valores de variables físicas se computa de alguna manera basándose en una descripción que está formulada en ese vocabulario [30].

Fodor tiene razón al decir que para tener algo parecido a un «problema psicológico de percepción» hemos de tener dicho modelo en la mente. Critica la sugerencia de Gibson de que podríamos evitar «el problema de cómo se detectan las (presuntas) invariantes estimulares» «distinguiendo entre el estímulo para los *transductores sensoriales* (es decir, las energías físicas) y el estímulo para los *órganos perceptivos* (es decir, invariantes abstractas)» diciendo:

> ... eso es incurrir en trivialización. Si se permite utilizar la idea de estímulo para distinguir el input a la retina (energía lumínica) del input al sistema óptico (patrones de energía lumínica que muestran invarianzas relevantes, por ejemplo, para la explicación de las constancias perceptivas), ¿por qué no hablar también del estímulo para *todo el organismo* (es decir, los perceptibles)? Así pues, la respuesta a «¿Cómo percibimos las botellas?» sería: «Es necesario y suficiente para la percepción de una botella que se detecte la presencia del estímulo invariante *botella.*»
>
> ... Lo que demuestra esto, pienso yo, no es que el problema psicológico de la percepción sea un embrollo, sino que la *formulación* del problema exige elegir (y motivar) un vocabulario propio para la representación de inputs. Ya he señalado que el vocabulario de valores de parámetros físicos *es* apropia-

[30] Jerry Fodor, *The Language of Thought* (Nueva York, 1975), pág. 47.

> do si se hace la suposición plausible de que los transductores sensoriales detectan valores de parámetros físicos y que todo el conocimiento perceptivo está mediado por la actividad de los transductores sensoriales (pág. 49n).

Aquí, Fodor aborda la cuestión que he planteado al examinar el intento de Quine de considerar la psicología como epistemología naturalizada: Si la elección de lo que tiene carácter de dato depende de algo más profundo que el consenso entre los observadores, expresable en el lenguaje corriente, ¿qué criterio puede seguir entonces el psicólogo al aislar el «input a nuestro mecanismo cognitivo»? Quine vacila en este punto, pero Fodor nos dice con firmeza y sensatez que, a no ser que consideremos como su input algo de lo que nuestro sujeto no necesite tener conocimiento, trivializaremos la idea de «procesamiento en el mecanismo cognitivo». A la pregunta retórica que he planteado antes, «¿Podría descubrir la psicología que el input al mecanismo cognitivo no está en la retina, sino más bien a mitad de camino en un lugar del nervio óptico?», Fodor respondería probablemenet: Sí, pues el hacerlo de una forma y no de otra depende de cuál es la forma de trazar líneas alrededor de las cajas negras que mejor divide el organismo en los trasductores y procesadores cuyas descripciones constituyen una teoría general y provechosa del procesamiento cognitivo.

Téngase en cuenta que esta respuesta elimina cualquier conexión entre la pregunta «¿Cómo reconocemos las botellas?» y la pregunta «¿Qué es lo que se da indubitablemente a la mente, de forma que pueda servir como piedra de toque infalible para la inferencia?» Efectivamente, la pregunta «¿Qué tiene derecho a creer el sujeto sin inferencia consciente?» o, más precisamente, «¿Qué clase de cosas puede justificar meramente con observaciones como 'Lo vi tan claramente como te veo a ti ahora' o 'Sé inglés'?» no tiene absolutamente nada que ver con la pregunta «¿Qué parte del organismo elegiremos como superficie de contacto con el mundo?» o, más exactamente, «¿Qué elección haremos de 'un vocabulario propio para la representación de inputs'?». Fodor es admirablemente explícito al respecto:

> Pero cualquiera que sea la relevancia que pueda tener para *algunos* objetivos la distinción entre estados del organismo y estados de su sistema nervioso, no hay ninguna razón particular para suponer que es relevante para los objetivos de la psicología cognitiva (pág. 52).
>
> ... los estados del organismo postulados en las teorías de la

cognición no figurarían como estados del organismo tratándose, por ejemplo, de una teoría de la responsabilidad legal o moral. ¿Y qué? Lo que importa es que deberían contar como estados del organismo para *algún* propósito útil. En especial, lo que importa es que deberían contar como estados del organismo en orden a construir teorías psicológicas que sean verdaderas (página 53),

Sólo nos queda añadir que lo que sirve para la responsabilidad moral o legal sirve también para la responsabilidad epistémica —que el organismo esté justificado al creer esto o aquello. No hay caminos que lleven del descubrimiento de las diversas superficies de contacto del organismo con el mundo a críticas de las opiniones del organismo sobre el mundo, o, en un plano más general, de la psicología a la epistemología. Lo que los empiristas supieron captar acertadamente al hablar de la percepción era que los órganos sensoriales deben considerarse dotados de un vocabulario que se empobrece al compararlo con el «vocabulario en que se formulan las hipótesis» —formulación realizada por la(s) unidad(es) de procesamiento o el propio sujeto. Tenían también razón, indudablemente, al felicitar a Galileo por preferir sus ojos a su Aristóteles, pero este juicio epistemológico no tiene ninguna conexión especial con su teoría de la percepción.

Ahora podemos ver cómo la imagen de la mente de Fodor, como un sistema de representaciones internas, no tiene nada que ver con la imagen del Espejo de la Naturaleza que vengo criticando. Lo más importante es que no hay forma de plantear la pregunta escéptica «¿Con qué perfección representan la realidad las representaciones internas del sujeto?» a su propósito del «lenguaje del pensamiento» de Fodor. En especial no es posible preguntar dónde, o con qué perfección, representan los productos de las teorías de la espontaneidad la fuente de la evidencia de la receptividad, y, por tanto, no hay posibilidad de ser escéptico sobre la relación entre apariencia y realidad. Tampoco hay nada general que decir aquí sobre el abismo que separa la evidencia presentada por la receptividad y las teorías creadas por la espontaneidad. Un abismo diferente se salva gracias al requisito de que el vocabulario de un mecanismo sensorial esté compuesto de «términos de clase natural de alguna ciencia física («teóricamente acabada)» (pág. 45). Este vocabulario estará «más empobrecido» que el de los «procesadores» en el sentido de que habrá una relación muchos-uno entre conjuntos de valores de parámetros físicos y los términos que utilizan los procesadores en sus hipótesis. Por eso, la relación será como la que existe entre ciencia y sentido común. La «infradeterminación de la teoría por la eviden-

cia», de Quine, queda así incorporada al modelo, en cuanto que muchas formas de hablar de sentido común serán compatibles con la única descripción verdadera (teóricamente terminada) del mundo por la ciencia física, y entre ellas pueden mediar muchos posibles lenguajes «procesadores». Sólo sería posible averiguar algo epistemológico aprendiendo los rasgos del «lenguaje del pensamiento» si, entre el vocabulario científico usado por los mecanismos sensoriales y los distintos vocabularios utilizados conscientemente por el sujeto, hubiera un vocabulario utilizado por los procesadores, cuyo conocimiento ayudaría al sujeto a descubrir la verdad sobre las cosas en general. Sólo esta clase de «realidad psicológica» podría ocupar el lugar de «lo que se da a la mente sin interpretación», como piedra de toque epistemológica de la justificación. Pero no es fácil ver por qué la realidad psicológica de ciertos colores, gramáticas o principios morales se va a corresponder con su uso al explicar o juzgar las realidades no psicológicas. Sólo esperaríamos que se diera esta correspondencia si consideráramos las unidades de procesamiento que aisla la teoría psicológica como, en cierto sentido, «lo mejor de nosotros», las sucesoras de la razón en cuanto soberana natural sobre el resto del alma, o como nuestro verdadero yo. Fodor y la psicología cognitiva no tienen ningún interés por conseguir esta situación honorífica para el código interno; ese código, como el FORTRAN, o la representación binaria de los números es *solamente* un código, no una ayuda para distinguir lo verdadero de lo falso.

Para ver este punto teniendo como trasfondo los capítulos anteriores, recordemos la afirmación de que la tradición epistemológica confundió las explicaciones causales de la adquisición de la creencia con las justificaciones de la creencia. Cuando las explicaciones causales se ofrecen en el código interno, la suposición de que se puede utilizar este código para detectar la adquisición de una creencia *verdadera* equivaldría a suponer que «los procesos mentales engendradores de verdad» eran clases naturales dentro de la teoría psicológica. Pero Fodor estaría probablemente de acuerdo en que no hay ninguna razón para pensar que estos términos evaluativos delimiten tales clases naturales. Como dice hablando de la creatividad:

> Puede ser que los procesos que consideramos creativos no formen una clase natural en el contexto de la explicación psicológica, sino que, por el contrario, cada *instancia* de dicho proceso sea una instancia de actividad dirigida por reglas y computacional de una u otra clase... Las categorías *creativo/aburrido* quizá sólo sean una clasificación cruzada con respecto a la taxonomía que utiliza la psicología.

> Sin embargo, lo que considero más importante es que el mero hecho de que los procesos mentales creativos sean procesos *mentales* no es garantía de que tengan una explicación en el lenguaje de la psicología bajo *cualquiera* de sus descripciones... Es posible que las buenas ideas... sean especies de estados mentales que no tienen causas mentales (págs. 201-202).

El ascenso de la idea de conocimiento como cuestión de representaciones internas debidamente ordenadas —un Espejo de la Naturaleza que ni se empaña ni distorsiona —fue debido a la idea de que la diferencia entre el hombre cuyas creencias eran verdaderas y el hombre cuyas creencias eran falsas dependía de «cómo funcionaban sus mentes». Si se toma esta expresión en el sentido de lo que dirían en una conversación», resulta verdadera, pero superficial y no filosófica. Para darle profundidad y carácter filosófico, hay que creer, con Descartes y Locke, que una taxonomía de entidades y procesos mentales llevará a descubrimientos que nos proporcionarán un método para descubrir la verdad, y no simplemente la verdad sobre la mente[31]. Pero la taxonomía psicológica considerada por Fodor no es una taxonomía epistemológica. Deja que el método y sustancia de las otras disciplinas que constituyen la cultura se hundan o se mantengan a flote por sus propias medios. Sólo la suposición de que algún día las distintas taxonomías elaboradas, por ejemplo, por Chomsky, Piaget, Lévi-Strauss, Marx y Freud formarán un solo cauce y deletrearán un grandioso Lenguaje Universal de la Naturaleza —suposición atribuida en ocasiones al estructuralismo— sugeriría que la psicología cognitiva tenía importancia epistemológica. Pero dicha sugerencia seguiría estando tan descaminada como la sugerencia de que, puesto que podemos prever cualquier cosa si tenemos el conocimiento suficiente de la materia en movimiento, una neurofisiología terminada nos ayudará a demostrar la superioridad de Galileo sobre sus contemporáneos. El abismo entre explicarnos a nosotros mismos y justificarnos a nosotros mismos es igual

---

[31] Cfr. Hiram Caton, *The Origins of Subjectivity: An Essay on Descartes* (New Haven, 1973), pág. 53: «La gran diferencia entre la metodología aristotélica y cartesiana es que, para Descartes, la mente es un principio de ciencia.» Contrapóngase la falta de conexión entre *De Anima* y *Analítica Posterior* con la suposición de Locke de que una investigación de «las bases y grados de creencia, opinión y asentimiento» se puede realizar con un «método histórico, sencillo» que comience con «el original de las ideas, nociones o comoquiera que se llamen, que el hombre observa y es consciente ante sí mismo de que tiene en la mente, y las formas en que el entendimiento se va equipando con ellas» (Essay, I, 1, ii-iii).

de grande tanto si en las explicaciones se utiliza un lenguaje de programación o un lenguaje de hardware.

Sin embargo, podría pensarse que si interpretamos la epistemología no como una forma de certificar nuestro éxito en el descubrimiento de la *verdad* sino como la creación de cánones de *racionalidad,* entonces un conocimiento del código interno nos ofrecerá algo para seguir adelante. Fodor, quizá en forma no intencionada, insinúa esta concepción cuando habla del descubrimiento de su código como demostración de «cómo está estructurada la racionalidad». Pero el único contenido que da a esta noción de «racionalidad» es el que presenta el siguiente pasaje:

> Si la línea argumental central de este libro es correcta, el lenguaje del pensamiento constituye el medio para representar internamente los aspectos psicológicamente destacados del entorno del organismo; en la medida en que es posible determinarlo en este lenguaje —y sólo en esa medida— esta información cae dentro de las rutinas computacionales que constituyen el repertorio cognitivo del organismo... Pero quisiera añadir algo más: que algunos organismos, al menos, parecen tener una libertad considerable para determinar cómo se va a utilizar este sistema representacional, y que dicha libertad suele explotarse generalmente de forma racional... Si los sujetos *calculan* realmente cómo se deben desplegar las representaciones internas, estos cálculos deben definirse, en relación con las representaciones; es decir, en relación con las representaciones de las representaciones. En resumen, ciertas propiedades del lenguaje deben estar representadas en el lenguaje del pensamiento, pues la capacidad de representar representaciones constituye, probablemente, una precondición de la capacidad de manipular representaciones de forma racional [32].

En este caso «racionalidad» significa la adaptación de medios y fines, y la capacidad de los organismos para hacer esto con sus representaciones difiere de su capacidad para hacerlo con sus hormonas únicamente en que para describir la primera capacidad hace falta un vocabulario metalingüístico. Pero comprender el vocabulario metalingüístico que utiliza el organismo para este objetivo no es comprender algo tan general como podría hacer pensar la expresión «la estructura de la racionalidad», sino más bien algo tan particular como los trucos que utiliza el programador para cerciorarse de que el computador pase de una subrutina a otra en la forma necesaria

[32] Fodor, *The Language of Thought,* pág. 172.

para optimizar la eficiencia. No entendemos más sobre lo que significa ser un agente o investigador racional por el hecho de comprender tales trucos que por el de entender qué es lo que hace que la glándula pituitaria libere esta hormona y no otra. Tampoco tendría sentido criticar la representación *consciente* que el sujeto tiene de su entorno —es decir, el vocabulario en que formula sus concepciones— argumentando que no representa estos aspectos tan bien como lo hacen «las rutinas computacionales que constituyen el repertorio cognitivo del organismo». «Racional» no vale más que «verdadero» (u «honrado» o «casto» o «bueno») como candidato a una idea evaluativa que pudiéramos entender mejor conociendo cómo funciona nuestra mente. Nuestro juicio sobre la racionalidad con que nos ha diseñado la evolución, o sobre lo racionales que ha conseguido hacernos la evolución, debe hacerse por referencia a nuestras concepciones sobre los fines que debemos servir. El conocimiento de cómo funciona nuestra mente no es más relevante que el conocimiento de cómo funcionan nuestras glándulas o moléculas para el desarrollo o corrección de dichas concepciones.

Si vamos a encontrar la relevancia epistemológica en la doctrina de las representaciones internas, tendrá que ser en las insinuaciones racionalistas de las opiniones «innatistas» comunes a Chomsky y a Fodor más que en su intento explícitamente antirreduccionista. Vendler nos presenta una inferencia de la idea de un lenguaje (y un metalenguaje) conectado, de Chomsky-Fodor, a una epistemología racionalista. Véase este pasaje, provocativamente anti-wittgensteiniano:

> ... la explicación más razonable es que el niño debe aprender su lengua nativa en forma parecida a como aprende una segunda lengua. En otras palabras, debe tener un equipamiento nativo que codifique los rasgos ilocucionarios, sintácticos y semánticos fundamentales de cualquier lenguaje humano posible... Este sistema de «ideas» nativas constituye el marco de referencia que luego se va rellenando progresivamente gracias a la influencia de un código más específico que representa los rasgos de la lengua materna... En cuanto al contenido de la reserva nativa de conceptos, lo único que podemos conseguir en la actualidad es hacer cálculos razonables. Sin embargo, en mi opinión, la tarea de exponerlo con todo detalle no es del todo imposible: Aristóteles, Descartes, Kant y, recientemente, Chomsky han conseguido delimitar dominios que deben pertenecer a este marco de referencia... Estas son, pues, las ideas «claras y distintas» que prestan inteligibilidad al resto. Son *a priori* en su origen y autónomas en su desarrollo: la experiencia no pue-

de cambiar su contenido. Ninguna experiencia es relevante para la propia idea de lo que es afirmar o pedir algo; para lo que es creer o decidir; para lo que es verdad o necesidad; para lo que es una persona, un objeto, un proceso o un estado; para lo que son el cambio, objetivo, causalidad, tiempo, extensión y número. Si estas ideas necesitan clarificación, la forma de conseguirla es reflexionar sobre lo que todos sabemos implícitamente y manifestamos en el uso correcto del lenguaje... [33]

Esta inferencia desde un vocabulario conectado hasta un conjunto de creencias que sólo puede ser «clarificado», no cambiado, tropieza con la crítica de Quine de las distinciones entre hecho y lenguaje, ciencia y filosofía, clarificar significados y cambiar creencias. Pero es más fundamental la objeción de que Vendler exige no sólo la premisa de que hay un lenguaje fijo del pensamiento sino también la premisa de que nuestro conocimiento de la naturaleza de ese lenguaje es inmune a la corrección basada en la experiencia. Esta es la misma premisa que utilizó Kant al explicar que podríamos entender nuestra posesión de verdades sintéticas a priori si y sólo si nuestra mente aportaba esas verdades [34]. Pero la afirmación de Fodor de que el descubrimiento del lenguaje del pensamiento es un proceso empírico muy prolongado tiene el corolario de que podemos equivocarnos siempre al decir qué es este lenguaje, y, por tanto, al determinar qué es a priori. La afirmación de Kant de que si sabemos lo que ocurre en el interior podemos legitimar nuestras certezas ante el tribunal de la razón pura recuerda la afirmación de Descartes de que «nada conoce la mente con tanta facilidad como a sí misma». Pero aunque la epistemología sea una disciplina más bien teórica, la psicología no lo es; ésta es una de las razones por las que la psicología no puede tener fines epistemológicos [35].

[33] Zeno Vendler, *Res Cogitans* (Ithaca, N. Y., 1972), págs. 140-141.

[34] K.d.r.V., Bxvii.

[35] Un intento no-chomskiano de vincular epistemología y psicología es el que nos ofrece Gilbert Harman en *Thought* (Princeton, 1973). Él encuentra la conexión entre ambas en el tema de los «ejemplos Gettier» —ejemplos en que la creencia justificada verdadera no es conocimiento porque, dicho en términos generales, se supone que una persona ha utilizado una falsa premisa en la inferencia que le llevó a la creencia en cuestión. Harman necesita una teoría de las «razones reales» para creer, y esto le lleva a lo que describe como «psicologismo» (págs. 15 y ss.). Sin embargo, no queda claro si Harman puede encontrar una conexión entre la investigación psicológica empírica y la postulación teórica de inferencias inconscientes específicas (que implican «razones reales») tal como exigen nuestras intuiciones sobre los ejemplos Gettier. Véase Michael Williams, «Inference, Justification and the Analysis of Knowledge», *Journal of Philosophy,* 75 (1978), 249-263, y Harman, «Using Intuitions

Puedo resumir este tema de las representaciones internas recurriendo una vez más a la confusión entre explicación y justificación. La idea de «representación», tal como la emplean los psicólogos, es ambigua, como si estuviera en una situación intermedia entre imágenes y proposiciones —entre, por ejemplo, las imágenes de la retina (o sus equivalentes en alguna parte de las profundidades de la corteza visual) y creencias como «Eso es rojo y rectangular». Sólo las últimas sirven de premisas, pero sólo las primeras son «no-mediadas», y la tradición del empirismo inglés las unió, con los resultados conocidos. Las representaciones-en-los-procesadores de Fodor son más proposiciones que imágenes, por lo que caen dentro de las críticas de Green y Sellars contra la idea empirista de «lo dado». Por otra parte, no son necesariamente proposiciones hacia las que el sujeto tenga actitudes. En realidad, las actitudes del sujeto hacia aquellas proposiciones de las que tiene conocimiento flotan con entera libertad en relación con las opiniones de los procesadores. Como señala Dennett en su crítica a Fodor, dos sujetos pueden tener la misma creencia aun cuando sus respectivos procesadores no hablen el mismo lenguaje [36]. Por eso, no debe haber ninguna *inferencia* que parta de las proposiciones mantenidas por los procesadores para llegar a las mantenidas por el sujeto, aun cuando la atribución de actitudes proposicionales diversas a los procesadores pueda ser la mejor forma posible de *explicar* cómo llegó el sujeto a mantener una creencia. A diferencia de las «ideas» del empirista, el proceso causal que va desde las imágenes de la retina, atraviesa diversas actitudes proposicionales mantenidas por varios procesadores para llegar a la producción del centro de habla del sujeto, no debe corresponderse necesariamente con ninguna cadena inferencial que justifique las concepciones del sujeto. La explicación puede ser privada, en el sentido de que, al menos por lo que podemos saber o nos interesa, los caprichos de la psicología podrían hacer que los amarillos o los peli-

---

about Reasoning to Study Reasoning: A Reply to Williams», *ibíd.,* 433-438. Si Harman es capaz de establecer este vínculo habrá aislado algo acertado en el intento de Locke de tratar el conocimiento en términos de los mecanismos del espacio interior. Pero todavía seguiría faltando un vínculo entre *justificación* y procesos psicológicos.

[36] Daniel Dennett, «Critical Notice» of *The Language of Thought, Mind,* 86 (1977), 278: «Si alguien está de acuerdo con Fodor en que es tarea de la psicología cognitiva trazar los procesos psicológicamente reales que se dan en las personas, entonces, dado que la adscripción de la creencia y el deseo sólo tiene con tales procesos una vinculación indirecta, podría decirse sin problemas que las creencias y deseos no son el objeto adecuado de estudio de la psicología cognitiva».

rrojos procesaran la información en lenguajes totalmente diferentes y con métodos distintos a los de los blancos, o de los que tienen pies planos. Pero la justificación es pública, en el sentido de que la discusión entre estas diferentes personas sobre qué hay que creer no hará probablemente ninguna referencia a cómo funcionan sus mentes caprichosas, ni debería hacerla. Por tanto, la afirmación de que poseemos un sistema de representaciones internas materializa, en el peor de los casos, no sólo la confusión entre imágenes y proposiciones sino también la confusión, más general, entre causalidad e inferencia.

Sin embargo, de hecho esta confusión sólo se da en las interpretaciones filosóficas de la psicología cognitiva, no en la explicación psicológica misma. Cuando los wittgensteinianos critican a la psicología, su blanco no es realmente la psicología sino la confusión de la epistemología con la psicología. Los psicólogos, por un deseo descaminado de ser «filosóficos», cometen a veces esta confusión. A los psicólogos contemporáneos, en protesta contra el conductismo, les gusta a veces considerarse a sí mismos como si estuvieran haciendo «científicamente» lo que Locke y Kant hicieron con lucubraciones. Pero hay una diferencia abismal entre decir:

> Debemos aislar los elementos no proposicionales de la conciencia que son el fundamento de la creencia en las proposiciones

y decir:

> Podemos tratar a *ítems* como los patrones de excitación neural como si fueran creencias para utilizar la metáfora de la «inferencia a partir de unos datos» al construir modelos de los procesos mentales [37].

Los psicólogos sólo tienen que afirmar lo segundo. Si se limitan a esto, pueden seguir a Putnam al tratar la distinción entre «proceso cerebral» y «proceso mental» como carente de interés filosófico superior al de la distinción entre «descripción del hardware» y «descripción del programa» [38]. La tentación de decir lo primero —la

---

[37] Sobre esta diferencia, véanse las críticas de J. O. Urmson, «Recognition», *Proceedings of the Aristotelian Society,* 56 (1955-1956), 259-280.

[38] Véase el artículo de Putnam «Minds and Machines», publicado en 1960 y reimpreso en *Mind, Language and Reality* (Cambridge, 1975), págs. 362-385 (especialmente los párrafos finales). Putnam fue el primer filósofo en señalar claramente que la conclusión de las analogías entre computadores y personas

tentación epistemológicamente motivada de «descubrir la conexión entre la mente y el cuerpo»— puede considerarse semejante a la tentación de plantear la pregunta «¿Cómo puede saber el computador que el patrón de cargas eléctricas que atraviesan el cable es el total de los ingresos monetarios del día?». Toda la concepción, surgida en el siglo XVII, según la cual podemos aprender más sobre lo que debemos creer entendiendo mejor cómo funcionamos, puede considerarse tan equivocada como la idea de que aprenderemos si hay que conceder derechos civiles a los robots comprendiendo mejor cómo funcionan. La analogía hombre-máquina se considerará útil no sólo como fuente de modelos provechosos de los organismos, sino también para ayudarnos a tener presente la diferencia entre los hombres como objetos de explicación y los hombres como agentes morales, interesados en justificar sus creencias y sus acciones. También puede ayudarnos, como expondré en los capítulos séptimo y octavo, a abandonar la idea de que hay que «sintetizar» estas dos formas de vernos a nosotros mismos.

---

no era que «los computadores nos ayudan a comprender la relación entre la mente y el cuerpo», sino más bien que «no puede haber ningún problema con la relación entre mente y cuerpo».

CAPÍTULO VI

Epistemología y filosofía del lenguaje

1. FILOSOFÍA DEL LENGUAJE PURA E IMPURA

La disciplina que se conoce actualmente con el nombre de «filosofía del lenguaje» tiene dos fuentes. Una es el conjunto de problemas apuntados por Frege y examinados, por ejemplo, por Wittgenstein en el *Tractatus* y por Carnap en *Meaning and Necessity*. Estos problemas se refieren a la manera de sistematizar nuestras ideas de significado y referencia de tal manera que se puedan aprovechar los progresos de la lógica cuantificacional, mantener nuestras intuiciones sobre la modalidad, y producir en general una imagen clara e intuitivamente satisfactoria de la forma en que encajan ideas como «verdad», «significado», «necesidad» y «nombre». A este conjunto de problemas le daré el nombre de objeto material de la filosofía «pura» del lenguaje —disciplina que no tiene ningún *parti pris* epistemológico, ni, de hecho, ninguna relevancia para la mayoría de los intereses tradicionales de la filosofía moderna. Puede rastrearse un antecesor de algunos de los problemas de Frege en Parménides, el *Sofista* de Platón, y algunas otras obras antiguas y medievales, pero son cuestiones que casi nunca se cruzan con otros «problemas de filosofía» de los manuales [1].

La segunda fuente de la filosofía del lenguaje contemporánea es explícitamente epistemológica. La fuente de esta filosofía del lenguaje «impura» es el intento de mantener la imagen de la filosofía de Kant por considerar que ofrece un marco permanente y ahistórico de la investigación en forma de una teoría del conocimiento. El «giro

[1] Piénsese en la dificultad de unir *El sofista* con *La república* o *Meaning and Necessity* con *Logical Structure of the World*.

lingüístico», como he dicho en el capítulo cuarto, comenzó como un intento por conseguir un empirismo no-psicologista reformulando las cuestiones filosóficas en forma de cuestiones de «lógica». Las doctrinas empiristas y fenomenalistas podrían considerarse ahora, se pensaba, como resultados del «análisis lógico del lenguaje» y no como generalizaciones psicológicas empíricas. En un sentido más general, las afirmaciones filosóficas sobre la naturaleza y alcance del conocimiento humano (por ejemplo, las que hizo Kant sobre las pretensiones de conocimiento en relación con Dios, la libertad y la inmortalidad) podrían formularse como observaciones sobre el lenguaje.

El tratamiento de la filosofía como análisis del lenguaje parecía unir los méritos de Hume con los de Kant. El empirismo de Hume parecía sustancialmente verdadero, pero con poca solidez metodológica pues no tenía otra base que una teoría empírica de la adquisición del conocimiento. Las críticas de Kant a la «mala» filosofía (por ejemplo, la teología natural) parecían al mismo tiempo más sistemáticas y más contundentes que las de Hume, pero parecían presuponer la posibilidad de una metodología no empírica. El lenguaje, a diferencia de la síntesis transcendental, parecía un campo de investigación convenientemente «natural» —pero, a diferencia de la psicología introspectiva, el análisis lingüístico parecía ofrecer la promesa de una verdad a priori. El decir que una sustancia material estaba constituida por la síntesis de una intuición diversa bajo un concepto a priori parecía «metafísico» mientras que decir que cualquier observación significativa sobre dicha sustancia podría formularse en términos de afirmaciones hipotéticas fenomenalistas parecía ser al mismo tiempo necesariamente verdadero y metodológicamente carente de misterio [2]. Kant había enseñado que la única forma en que podría ser posible el conocimiento a priori sería que fuera conocimiento de nuestra aportación —aportación de nuestra facultad de espontaneidad— a la constitución del objeto del conocimiento. Reformulado por Bertrand Russell y C .I. Lewis, esto se convirtió en la concepción de que toda afirmación verdadera contenía nuestra aportación (en forma de significados de los términos componentes) así como la del mundo (en forma de los hechos de la percepción sensorial).

---

[2] Cfr. Hilary Putnam, *Mind, Language and Reality* (Cambridge, 1975), páginas 14-19, donde estudia el idealismo y el fenomenalismo. Putnam presenta la idea tradicional, que yo vengo rechazando, de que el «giro lingüístico» permitió a los filósofos presentar soluciones positivas a problemas tradicionales.

El ataque contra esta última idea, que he descrito en el capítulo cuarto, ha producido dos movimientos curiosamente diferentes en la reciente filosofía del lenguaje. Uno de ellos adquiere su mejor representación en Davidson, y el otro en Putnam. La primera reacción es en dirección a una concepción purificada y des-epistemologizada de la filosofía del lenguaje. Un resultado de esta remodelación del tema es el rechazo de lo que Davidson llama «el tercer dogma» del empirismo, a saber, el «dualismo de esquema y contenido, de sistema organizador y algo que está a la espera de ser organizado» —dogma que, como he demostrado en el capítulo cuarto, es fundamental para la epistemología en general así como para el empirismo en particular [3]. Davidson distingue entre proyectos filosóficos que forman parte de «la teoría del significado propiamente dicha» y los que están motivados por «cierto puritanismo filosófico adventicio» [4]. En términos generales, Frege y Tarski trataron de realizar el primer tipo de proyecto, mientras que Russell y Carnap y Quine mezclaron la teoría pura del significado con consideraciones epistemológicas impuras —las que les llevaron, en varias ocasiones y maneras, a formas diversas de operacionalismo, verificacionismo, conductismo, convencionalismo y reduccionismo [5]. Cada uno de éstos era una expresión de un «puritanismo filosófico» subyacente que afirmaba que todo lo que fuera incapaz de ser «construido lógicamente» a partir de certezas (los datos de los sentidos, o las normas del lenguaje) resultaba sospechoso.

En opinión de Davidson, la cuestión de «cómo funciona el lenguaje» no tiene ninguna conexión especial con la cuestión de «cómo funciona el conocimiento». El hecho de que se considere la verdad en conexión con ambas cuestiones no debería engañarnos y llevarnos a pensar que podemos deducir:

> Una teoría del significado analizará el significado de todas las expresiones referentes que no sean las de los *qualia* sensoriales en términos de aquellas que sí se refieren a los *qualia* sensoriales

a partir de

[3] Véase Donald Davidson, «On the Very Idea of a Conceptual Scheme», *Proceedings of the American Philosophical Association,* 17 (1973-74), 11.

[4] Donald Davidson, «Truth and Meaning», *Synthese,* 7 (1967), 316.

[5] Putnam critica mordazmente a Quine en este punto. Cfr. Putnam, *Mind, Language and Reality,* págs. 153-191 («The Refutation of Conventionalism»). Sin embargo, como veremos, el propio Putnam incurre en esta misma confusión.

> Nuestra única evidencia en favor de las verdades empíricas son los patrones de los *qualia* en nuestros campos sensoriales.

Para Davidson, una teoría del significado no es una acumulación de «análisis» de los significados de los términos individuales, sino más bien una comprensión de las relaciones inferenciales entre las oraciones [6]. Entender estas relaciones es entender las condiciones de verdad para las oraciones en inglés, pero para muchas oraciones simples («Un roble es un árbol», «Rusia es nuestra patria», «La muerte es inevitable») no se pueden presentar condiciones de verdad más instructivas que las presentadas para «La nieve es blanca».

El caso es distinto con oraciones que adscriben creencias o acciones, o con las que contienen modificadores adverbiales —o cualquier otra oración tal que las relaciones inferenciales que se obtengan entre ella y las oraciones próximas no se revelen por el aparato ordinario de la lógica cuantificacional sin un nuevo análisis gramatical. En estos casos, conseguimos condiciones de verdad que no son triviales, son difíciles de construir, sólo se pueden comprobar mediante su susceptibilidad a integrarse dentro de una teoría de las condiciones de verdad para otras oraciones. «El efecto deseado», dice Davidson, «es clásico en la construcción de teorías: extraer un concepto rico (en este caso algo razonablemente parecido a la traducción) de fragmentos pequeñitos de evidencia (en este caso los valores de verdad de las oraciones) imponiendo una estructura formal a bastantes fragmentos» [7]. Este programa no sólo no tiene nada que ver con la epistemología, sino que sus resultados «ontológicos» no pueden tener mucha fuerza. Por ejemplo, no sirve a ninguno de los objetivos sentimentales habituales de la construcción de sistemas metafísicos el enterarse por Davidson de que probablemente tendremos que seguir hacia adelante y hacer cuantificaciones con las personas, en vez de «reducirlas» a cosas, para tener una teoría de la verdad para las oraciones de acción. En realidad el interés filosófico del programa es en gran parte negativo: al hacer ver a qué viene a parar la filosofía del lenguaje cuando se la purifica de los intentos de imitar a Kant o a Hume, hace que resalten con toda claridad los «puritanismos adventicios» de los programas anteriores. Los re-

---

[6] Véase Davidson, «Truth and Meaning», págs. 316-318.

[7] Donald Davidson, «In Defense of Convention T» en *Truth, Syntax and Modality*, ed. H. Leblanc (Amsterdam, 1973), pág. 84. Sobre el resultado moderado de la teoría, véase Davidson, «The Method of Truth in Metaphysics», *Midwest Studies in Philosophy*, 2 (1977), págs. 244-254, especialmente los párrafos finales.

sultados reales del trabajo intenso en la modificación adverbial y temas semejantes que se conseguirían con los esfuerzos concertados por realizar las sugerencias de Davidson no harían gran cosa para favorecer o entorpecer la solución de ninguno de los problemas de filosofía que aparecen en los manuales.

La mejor manera de entender la obra de Davidson es considerar que ha llevado a cabo la disolución de Quine de la distinción entre cuestiones de significado y cuestión de hecho —su ataque a la reinterpretación de la distinción kantiana entre la receptividad del sentido y los conceptos a priori presentados por la espontaneidad. Lo que dice Davidson es que obramos con seriedad al renunciar a un conocimiento a priori del significado, la teoría del significado será una teoría empírica. Por eso, no puede haber ninguna parcela especial para esta teoría, a no ser, más o menos, la parcela tradicional del gramático —el intento de encontrar formas de describir oraciones que ayuden a explicar cómo se utilizan esas oraciones. Desde esta perspectiva, la «notación canónica» de Quine no debe entenderse como un intento «de dibujar la estructura verdadera y última de la realidad»[8], sino más bien como un intento de encontrar la forma más clara de describir una parcela de la realidad relativamente pequeña —el uso del lenguaje. El sentido de la construcción de una «teoría de la verdad del inglés» no es hacer posible que los problemas filosóficos se puedan exponer en un modo de habla formal, ni explicar la relación entre palabras y mundo, sino simplemente trazar claramente la relación entre unas partes de una práctica social (el uso de ciertas oraciones) y otras partes (el uso de otras oraciones).

Dentro de la actual filosofía del lenguaje, el enfoque opuesto al de Davidson se encuentra representado por Dummett y Putnam. Dummett todavía se mantiene fiel al lema común a Viena y Oxford —la afirmación de que «la filosofía tiene, como primera aunque no única tarea, el análisis de los significados». Davidson, por lo que yo puedo entender, no tiene este apego a la idea de «analizar significados». Dummett continúa diciendo algo que no dice Davidson, y que, por lo que yo puedo entender, no tendría ninguna razón para decir: «la teoría del significado, que es la búsqueda de dicho modelo, es el fundamento de toda la filosofía, y no la epistemología, como Descartes nos hizo creer equivocadamente»[9]. Según el punto

---

[8] W. V. O. Quine, *Word and Object* (Cambridge, Mass., 1960), pág. 221.

[9] Michael Dummett, *Frege's Philosophy of Language* (Londres, 1973), página 559. En sus polémicas contra el holismo de Davidson, Dummett insiste en que no se puede tener una filosofía del lenguaje adecuada sin las dos

de vista que estoy adoptando en el presente libro, esta afirmación es engañosa en muchos sentidos. Es erróneo decir que Descartes nos hizo crer equivocadamente que la epistemología era el fundamento de toda la filosofía. Más bien, lo que hizo fue posibilitar que Locke y Kant elaboraran una problemática epistemológica que sustituyera a la problemática escolástica. Hizo posible una disciplina en la que la metafísica consistiera en hacer del mundo un lugar seguro para las ideas claras y distintas y para la obligación moral, y en la que los problemas de la filosofía moral se convirtieran en problemas de meta-ética, problemas de la justificación del juicio moral. Esto no es tanto hacer de la epistemología el fundamento de la filosofía cuanto inventar algo nuevo —la epistemología— que lleve el nombre de «filosofía». Resulta bastante plausible la imagen de que la filosofía antigua y medieval se ocupa de las *cosas,* la filosofía del siglo XVII al XIX de las *ideas,* y la filosofía contemporánea ilustrada de las *palabras.* Pero no debe pensarse que esta secuencia ofrezca tres opiniones contrarias sobre lo que es primario o lo que tiene carácter de fundamento. No es que Aristóteles pensara que la mejor forma de explicar las ideas y palabras era hacerlo en términos de las cosas, mientras que Descartes y Russell adaptaron de otra forma el orden de la explicación. Sería más correcto decir que Aristóteles no tenía —no sentía necesidad de— una teoría del conocimiento, y que Descartes y Locke no tenían una teoría del significado [10]. Las observaciones de Aristóteles sobre el conocer no ofrecen respuestas, buenas o malas, a las preguntas de Locke, lo mismo que las observaciones de Locke sobre el lenguaje no frecen respuestas a las de Frege.

Dummett piensa que la filosofía del lenguaje tiene carácter de fundamento porque ve que las cuestiones epistemológicas se están formulando ahora, por fin, como cuestiones dentro de la teoría del significado. Está de acuerdo con Descartes sobre la importancia de las cuestiones que surgieron del «camino de las ideas», pero piensa que sólo en época reciente hemos sido capaces de formularlas adecuadamente. Lo mismo que Spinoza y Leibniz pensaban que ellos estaban haciendo bien lo que los metafísicos escolásticos habían he-

distinciones kantianas (dado-versus-interpretado y necesario-versus-contingente) atacadas por Quine y Sellars. Véase, especialmente, «What Is a Theory of Meaning? (I)» en Samuel Guttemplan, ed., *Mind and Language* (Oxford, 1975), páginas 97-138. La defensa de la primera distinción se hace más explícitamente en la pág. 137 y la de la segunda en las págs. 117 y ss.

[10] Véase Ian Hacking, *Why Does Language Matter to Philosophy?* (Cambridge, 1975), pág. 43.

cho muy torpemente (por ejemplo, investigar la naturaleza de la sustancia), también Dummett y Putnam piensan que están haciendo bien lo que los epistemólogos han hecho con torpeza (investigar el problema del realismo e idealismo). Pero Descartes, como podemos ver retrospectivamente, se había engañado a sí mismo; para el tiempo de Kant estaba claro que si se partía de los problemas cartesianos no se podrían llegar a plantear las antiguas cuestiones metafísicas. La relación de Davidson con Russell, Putnam y Dummet, en este sentido, es la misma que había entre Kant y Descartes, Spinoza y Leibniz. La distinción de Davidson entre presentar una teoría del significado y el «puritanismo filosófico adventicio» es un paralelo contemporáneo de la distinción de Kant entre los usos de la razón legítimos y los ilegítimos.

Si adoptamos esta idea de que los paradigmas filosóficos nuevos van dejando de lado a los antiguos problemas, en vez de ofrecer nuevas formas de formularlos o resolverlos, consideraremos el segundo tipo de filosofía del lenguaje (el «impuro») como un último intento nostálgico de conectar un nuevo tipo de actividad filosófica con una problemática antigua. La idea de Dummett de la filosofía del lenguaje como «filosofía primera» nos parecerá equivocada no porque haya alguna otra área «primera» sino porque la idea de la filosofía en cuanto dotada de fundamentos es tan errónea como la de que el conocimiento tenga fundamento. Según esta concepción, «filosofía» no es el nombre para una disciplina que aborda cuestiones permanentes y, por desgracia, no llega a formularlas adecuadamente ó se limita a atacarlas con torpes instrumentos dialécticos. Es, más bien, un género cultural, una «voz en la conversación de la humanidad» (en palabras de Michael Oakeshott), que se centra en un determinado tema y no en otro en un momento dado, no por necesidad dialéctica sino como consecuencia de varias cosas que ocurren en otra parte de la conversación (la Nueva Ciencia, la Revolución Francesa, la novela moderna) o de hombres individuales con talento que piensan algo nuevo (Hegel, Marx, Frege, Freud, Wittgenstein, Heidegger), o quizá de la resultante de varias de estas fuerzas. Un cambio filosófico interesante (podríamos decir «el progreso filosófico», pero esto sería una petición de principio) se produce no cuando se encuentra una nueva forma de hacer frente a un problema antiguo sino cuando aparece un nuevo conjunto de problemas y los antiguos comienzan a esfumarse. Pero, por todas las razones que Kuhn y Feyerabend han presentado en su crítica al enfoque de la historia de la investigación que es típico de los «manuales», hay que resistir a esta tentación.

En este capítulo voy a tratar, en su mayor parte, de defender esta concepción «kuhniana» de la relación entre filosofía del lenguaje y problemas filosóficos tradicionales, en oposición a la forma de entender su relación, más conocida (y, en mi opinión, generalmente aceptada), propuesta por Dummett. Para ello haré una crítica de la filosofía «impura» del lenguaje y de sus ilícitas transferencias de problemas desde la epistemología a la filosofía del lenguaje. Evidentemente, no puedo estudiar todo este campo, por lo que me limitaré a un tema que constituye la mayor tentación de pensar que las explicaciones de cómo funciona el lenguaje nos ayudarán también a ver «qué conexión hay entre el lenguaje y el mundo» y, por tanto, cómo son posibles la verdad y el conocimiento. Se trata de la cuestión llamada del cambio conceptual, en torno a la cual se han centrado las polémicas más recientes sobre «realismo», «pragmatismo», «verificacionismo», «idealismo» y «convencionalismo». Creo que Davidson tiene razón al decir que la idea de «cambio conceptual» es en sí mismo incoherente, y que debemos profundizar en ella para reconocer por qué los debates en torno a ella parecen a la vez tan importantes y con tan pocas probabilidades de resolución.

Por eso, en la próxima sección, examinaré las cuestiones planteadas deliberadamente por el cambio de teoría en la epistemología, por una parte, y en la filosofía del lenguaje por la otra. Luego, en posteriores secciones, estudiaré con cierto detalle la respuesta «realista» de Putnam a estas cuestiones, pues la obra de Putnam es, en mi opinión, la formulación más clara de un programa «impuro». La compararé con un enfoque del lenguaje «puro» o «pragmatista» o «del juego lingüístico», que aparece reflejado en Sellars y en Wittgenstein, así como en Davidson (a pesar de las diferencias que puedan hacernos pensar que los tres adoptan posturas opuestas). Con esto espero demostrar que la cuestión de los dos enfoques no es una repetición de los enfrentamientos que separaban a los realistas de los idealistas y pragmatistas en los días de la filosofía-como-epistemología, y no es realmente, en absoluto, una cuestión lingüística. Lo que voy a defender es que si llevamos hasta más allá de cierto punto las críticas de Quine y Davidson a las distinciones lenguaje-hecho y esquema-contenido, nos quedamos sin espacio dialéctico para exponer el enfrentamiento entre el «realista» y el «idealista» (o «pragmatista») en relación con el tema de «qué conexión tiene el lenguaje con el mundo». La necesidad de plantear esta cuestión me parece una manifestación más de la necesidad kantiana de una matriz neutra, permanente, y englobadora, dentro de

la que se puedan «colocar» y criticar las investigaciones pasadas y futuras. Esta nostalgia por la filosofía en cuanto disciplina arquitectónica y envolvente sobrevive en la filosofía del lenguaje contemporáneo únicamente por obra de la vaga asociación del «lenguaje» con «lo a priori» y de esto último con la «filosofía». En la medida en que hay un enfrentamiento real entre «realistas» y «pragmatistas» en la filosofía del lenguaje contemporáneo me parece que es de carácter metafísico —la cuestión de si la filosofía puede mantener su auto-imagen kantiana una vez que se abandona la idea del lenguaje como fuente de conocimiento *a priori*. En *este* punto, sin duda ninguna, me pondré del lado de los «pragmatistas». Pero espero que mi exposición sirva para enmarcar este problema metafilosófico en un lenguaje menos misterioso que el de sus formulaciones habituales.

2. ¿De qué hablaban nuestros antepasados?

¿Se equivocaba Aristóteles cuando decía que el movimiento se dividía en natural y forzado? ¿O estaba hablando de algo diferente de lo que hablamos nosotros cuando hablamos de movimiento? ¿Ofreció Newton respuestas correctas a las preguntas a las que Aristóteles había ofrecido respuestas erróneas? ¿O se hacían preguntas diferentes? Este tipo de enigma ha inspirado una gran parte de los trabajos más interesantes realizados en la filosofía de la ciencia y en la filosofía del lenguaje durante los últimos años. Sin embargo, como la mayoría de los misterios filosóficos, sus motivos y presuposiciones tienen más interés que las distintas soluciones que se han presentado. En definitiva, ¿por qué debemos pensar que estas preguntas tienen una respuesta más interesante que la respuesta a la pregunta de si el barco de Teseo resistió el cambio de cada una de sus tablazones? ¿Por qué hemos de pensar que la pregunta «¿Qué querían decir?» o «¿A qué se estaban refiriendo?» va a tener una respuesta determinada? ¿Por qué va ser posible responder en uno u otro sentido, según cuáles sean las consideraciones heurísticas relevantes para un objetivo historiográfico concreto?

La razón por la que pensamos que debería haber una respuesta determinada es, en una primera aproximación, que pensamos que la historia de la búsqueda de la verdad debe ser diferente de la historia de la poesía o de la política o del vestido. Es muy posible que tengamos la impresión de que preguntas como «¿Querían los griegos decir 'templanza' cuando decían σωφροσύνη?» y «¿Se re-

fieren los Nuer al alma como *kwoth*?» se puedan despachar diciendo que no hay ninguna razón especial para pensar que cualquier expresión de una palabra de una cultura pueda emparejarse con una expresión de una sola palabra de una cultura diferente. En realidad, podemos tener la sensación de que ni siquiera con largas paráfrasis se conseguiría mucho, y de que hemos de introducirnos en el vaivén del exótico juego lingüístico [11]. Pero en el caso de la ciencia, esta actitud parece algo terca. Aquí, nos sentimos inclinados a decir, es claro que hay algo ahí fuera —el movimiento y sus leyes, por ejemplo— a lo que las personas trataban de referirse o, al menos, a lo que se estaban refiriendo sin darse cuenta. Se supone que la investigación científica descubre qué clases de objetos hay en el mundo y qué propiedades tienen. Quien realizara una investigación seria lo único que podría preguntarse era qué predicados había que asignar a cada cosa. Cuando encontramos dificultades para decir de qué hablaba Aristóteles, tenemos la sensación de que tiene que haber, en alguna parte, una respuesta correcta, pues tenía que estar hablando de *algunas* de las cosas de que hablamos *nosotros*. Aun cuando se imaginara objetos irreales y propiedades no realizadas en casos concretos, no podía por menos de dar sentido a lo que decía sobre ellas hablando en forma entrelazada de lo que había *realmente* o mediante una interacción con ello. Esta sensación constituye la raíz de afirmaciones instrumentalistas como «Tanto hablar de clases y leyes de movimiento no es más que una forma complicada de clasificar las experiencias sensoriales». Esta necesidad de decir que el hablar de algo que no reconocemos es «realmente» hablar de algo que sí reconocemos solía satisfacerse suponiendo sin más (al estilo «Whiggish») que nuestros desorientados antepasados habían estado hablando «realmente» de aquello que nuestros investigadores contemporáneos más respetados decían que habían estado hablando. Así oíamos decir que Aristóteles había estado hablando realmente de la gravitación cuando hablaba del movimiento descendente natural, que los ignorantes marinos se referían realmente a los cuernos del narval cuando hablaban de los del unicornio, que «el fluido calórico» era una forma equivocada de describir la transmisión de energía entre moléculas que rebotaban, y que Kierkegaard estaba describiendo nuestras relaciones con nuestros padres carnales cuando hablaba de la relación de Abraham con Dios.

En los últimos años ha habido dos cambios que han hecho po-

---

[11] Véase Clifford Geertz, «Thick Description: Toward an Interpretive Theory of Culture» en su *The Interpretation of Cultures* (Nueva York, 1973).

nerse nerviosos a los filósofos en relación con esta estrategia de «de lo que estaban hablando realmente». Estas innovaciones son, en términos generales, consecuencia del hundimiento de los dos «dogmas del empirismo» identificados por Quine. El primer dogma entronizaba lo que Quine llamó «esencialismo» —la idea de que se podía distinguir entre aquello de lo que hablaba la gente y lo que decían sobre ello, descubriendo la esencia del objeto en cuestión. En su forma lingüística era la doctrina de que se podía descubrir cuál era el término de nuestro idioma que traducía otro término en el de los científicos antiguos, y luego descubrir la esencia del referente de ambos términos distinguiendo entre las afirmaciones analíticas que decían el significado de los términos y las afirmaciones sintéticas que expresaban creencias, posiblemente falsas, sobre este referente. El segundo dogma afirmaba que esta traducción era algo que se *podía* encontrar siempre, y que siempre se *podían* formular estas afirmaciones analíticas, pues para determinar el significado de cualquier expresión referente sólo hace falta descubrir qué informes de un «lenguaje de observación neutral» se verían confirmados, y cuáles desconfirmados, por una declaración que afirmara la existencia del referente en cuestión.

La convicción de que la ciencia difería del discurso menos estricto por el hecho de tener «referencia objetiva» a las cosas «de ahí fuera» se vio reforzada antes de Quine por el pensamiento de que si no hubiera cosas que fueran como las esencias aristotélicas, que podían hacerse inmaterialmente presentes en el intelecto, había ciertamente puntos de contacto con el mundo en las presentaciones de los sentidos. Este contacto, más la capacidad de un «análisis del significado» operacionalista para describir la esencia del referente en términos de las presentaciones que se podían esperar de él, parecía dar a la ciencia algo de que carecían la religión y la política —la capacidad de usar el contacto con lo real como piedra de toque de la verdad—. El horror con que se recibió el derrocamiento de los dogmas conseguido por Quine, y los ejemplos de la «carga de teoría» de la observación propuestos por Kuhn y Feyerabend, fue consecuencia del temor de que no hubiera tal piedra de toque. Si llegamos a admitir que Newton era superior a Aristóteles no porque sus palabras se correspondieran mejor a la realidad sino sencillamente porque Newton nos permitía desenvolvernos mejor, ya no quedaría nada que permitiera distinguir la ciencia de la religión o de la política. Era como si todo lo que quedara entre nosotros y el «irracionalismo» fuera la capacidad de distinguir lo analítico de lo sintético, y lo observacional de lo teórico.

En capítulos anteriores he indicado que esta identificación de la racionalidad con los dogmas filosóficos de la época refleja el hecho de que, desde Kant, la filosofía se ha dedicado a presentar un marco neutro y permanente de la cultura. Este marco de referencia se construye en torno a una distinción entre la investigación de lo real —las disciplinas que están en «el camino seguro de una ciencia»— y el resto de la cultura. Esta es la clase de distinción que encontramos en el último párrafo de la *Enquiry* de Hume, en el primer párrafo del prefacio a la segunda edición de la *Crítica de la razón pura,* y en los manifiestos de Russell y Husserl citados en el capítulo cuarto. Si la filosofía es esencialmente la formulación de la distinción entre ciencia y no ciencia, entonces el poner en peligro las formulaciones actuales parece poner en peligro la filosofía misma, y con ella la racionalidad (de la que la filosofía sería el guardián vigilante, en constante lucha por rechazar a las fuerzas de las tinieblas). Dada esta concepción de los dogmas que Quine desbarató, la respuesta de muchos filósofos fue encontrar alguna manera de formular la deseada distinción que *a)* mantuviera a la filosofía del lenguaje en el mismo lugar central que había ocupado desde la época del Círculo de Viena, *b)* no implicara recurso ninguno a la idea del lenguaje como el mundo del *a priori, c)* proporcionara una respuesta a las preguntas sobre si Newton y Aristóteles tenían un referente común (y en caso afirmativo, cuál era). Este deseo fue la raíz de lo que ha venido a llamarse «la teoría de la referencia», término que ha llegado a ser coextensivo con lo que yo vengo llamando «filosofía del lenguaje impura».

Pero antes de que este deseo llegara a explicitarse, hubo una etapa preliminar de desconcierto en que los filósofos se pusieron a pedir una teoría del «cambio de significado». Esta demanda surgió principalmente como consecuencia de una reacción a la afirmación de Feyerabend de que la concepción empirista tradicional había presupuesto una «condición de invarianza del significado» —es decir, que «todas las teorías futuras deberán estar formuladas de tal manera que su uso en las explicaciones no afecte a lo que dicen las [otras] teorías, ni a los informes de hecho que se deban explicar» [12]. A Feyerabend, como a Kuhn, le interesaba demostrar que el significado de muchas afirmaciones del lenguaje, incluyendo muchas afirmaciones de «observación», cambiaban cuando aparecía una nueva teoría; o, al menos, que si se admitía la existencia de dicho cam-

[12] Paul Feyerabend, «How to Be a Good Empiricist» en *Challenges to Empiricism,* ed. Harold Morick (Belmont, Calif., 1972), pág. 169.

bio se podían explicar mucho mejor los hechos de la historia que con la concepción clásica de los manuales según la cual los significados *podían* cambiar como resultado de nuevos experimentos —que el marco neutral y permanente de significados dentro del cual se podía realizar la investigación racional no era tan permanente como se había pensado. Pero, decían ellos, debe haber un cambio de significado que sea «racional» y dirigido por unos principios, y lo que debemos hacer nosotros, como guardianes y explicadores de la racionalidad que es natural para los científicos de la naturaleza, es explicar cuáles son los principios que entran en juego. El propio Feyerabend se había limitado a señalar que los significados cambiaban cada vez que cambiaba un rasgo de uso, pero mentes más frías pensaban que debía haber un término medio entre «los significados permanecen y las creencias cambian» y «los significados cambian siempre que cambian las creencias». De ahí surgió la sensación de que tenía que haber una forma de distinguir un cambio de creencias dentro de la «ciencia normal» de Kuhn y el cambio de normas que ocurría en una «revolución científica». Aun admitiendo, decían los filósofos, que la explicación del cambio de teoría que aparece en los manuales sea engañosa, sin embargo, todo lo que necesite el historiógrafo de la ciencia se lo puede proporcionar la filosofía. Nosotros intentaremos descubrir aquellas condiciones en que cambios sucesivos de creencias producen algo que no es meramente un cambio de creencias sino un cambio de «esquema conceptual» [13].

Durante algún tiempo se dejó sentir la tentación de admitir que sería correcto hacer la igualdad de significado, la objetividad y la

[13] Fred Suppe, en su «The Search for Philosophic Understanding of Scientific Theories» (en *The Structure of Scientific Theories,* ed. Suppe [Urbana, 1974], páginas 3-241), clasifica este proyecto como un intento de construir un «análisis *Weltanschauungen*» del cambio de teoría científico (cfr. especialmente las págs. 127 y ss.). Muchos de los autores de quienes se ocupa Suppe se opondrían a las implicaciones del término, pienso yo, y a algunos de los detalles de la presentación de Suppe. Pero las líneas generales del tratamiento que hace Suppe de este periodo dentro del desarrollo de la filosofía reciente de la ciencia me parece exacto y revelador. Otro documento útil para entender este periodo es *Conceptual Change,* ed. Glenn Pearce y Patrick Maynard (Dordrecht, 1973), que tiene algunos artículos muy útiles sobre la cuestión del «cambio de significado», especialmente los de Binkle Sellars, Putnam, Barrett y Wilson. La línea que estoy siguiendo en este capítulo sintoniza con los artículos de Binkley y Sellars, pero creo que ambos se toman demasiado en serio la pregunta que Binkley formula así: «¿Cómo debe aplicarse nuestro sistema de valoración epistémica en contextos de significados cambiantes?» (pág. 71). En mi opinión, no existe tal sistema —ninguna estructura de la racionalidad que tenga alcance general. Las críticas de Barret a Putnam concuerdan con las presentadas por Fine y por mí en los artículos citados más adelante en la nota 26.

verdad relativas a un esquema conceptual, con tal que hubiera algunos criterios para saber dónde y por qué era racional adoptar un nuevo esquema conceptual. El filósofo, guardián de la racionalidad, se convirtió en el hombre que decía cuándo se podía comenzar a querer decir algo diferente, en vez de ser únicamente el hombre que decía lo que se quería decir. Pero este intento de conservar el papel tradicional del filósofo se vio condenado al fracaso. Todas las razones de Quine por las que no podía hacer una cosa eran también razones por las que no podía hacer la otra. El filósofo había aparecido, desde los primeros momentos del «giro lingüístico», como un hombre que sabía de conceptos porque sabía de los significados de las palabras, y cuya tarea había transcendido lo empírico. Pero en cuanto se admitió que «las consideraciones empíricas» (por ejemplo, el descubrimiento de que había manchas en la luna, el descubrimiento de que los *États-Généraux* no conseguirían sus objetivos) incitaban pero no imponían el «cambio conceptual» (por ejemplo, un concepto diferente de los cielos o del estado), ya no tenía sentido la división entre el trabajo del filósofo y el del historiador. En el momento en que se decía que era razonable abandonar el esquema conceptual aristotélico como consecuencia de algún descubrimiento, el «cambio de significado» o el «cambio de esquema conceptual» no significaban más que «cambio en creencias especialmente importantes». El historiador puede hacer inteligible el cambio del esquema antiguo al nuevo, y hacernos ver por qué habríamos pasado del uno al otro si hubiéramos sido un intelectual de aquella época. No hay nada que pueda añadir el filósofo a lo que ya ha hecho el historiador para demostrar que este curso inteligible y plausible es «racional». Sin lo que Feyerabend llamaba «invarianza del significado», no hay ningún método especial (análisis del significado) que pueda aplicar el filósofo. La «invarianza de significado» era simplemente la forma «lingüística» de formular la afirmación kantiana de que la investigación, para ser racional, debía realizarse dentro de un marco permanente cognoscible *a priori,* un esquema que restringiera el posible contenido empírico y explicara qué era más racional hacer con cualquier contenido empírico que se presentara. Una vez que los esquemas pasaron a ser temporales se puso en peligro la misma distinción esquema-contenido, y con ella la idea kantiana de la filosofía en cuanto posibilitada por nuestro conocimiento *a priori* de nuestra propia aportación a la investigación (el elemento esquemático, formal —por ejemplo, el «lenguaje»).

3. El idealismo

Los filósofos llegaron a darse cuenta bastante pronto de que la búsqueda de un criterio de cambio del significado resultaba tan desastrosa para la idea de la filosofía en cuanto análisis de significados como la idea del derecho a la rebelión lo había sido para la filosofía política en cuanto estudio de la soberanía. Así pues, se dieron cuenta de que Feyerabend se había expresado mal al hablar de «cambio del significado». Fue Putnam quien presentó esta crítica de forma más efectiva:

> Feyerabend no puede escapar de las mismas dificultades que han importunado a los Positivistas... Para ver que las cosas son así, baste recordar que para Feyerabend el significado de un término depende de *toda* una teoría que contiene el término... Como es lógico, alguien podría seguir la línea radical y decir que *cualquier* cambio en la teoría es un cambio en el significado de los términos... Pero supongo que Feyerabend no seguiría esa línea. Pues decir que cualquier cambio en nuestras creencias empíricas sobre Xs es un cambio en el significado del término X sería abandonar la distinción entre cuestiones de significado y cuestiones de hecho. Decir de que las reglas semánticas del inglés no pueden distinguirse en absoluto de las creencias empíricas de los que hablan inglés sería tirar por la borda la noción de una regla semántica del inglés... Si Feyerabend hubiera seguido este curso, habría desaparecido toda apariencia de sensación. Porque en este contexto la «sensación» depende de la realización de movimientos de avance y retroceso entre una concepción no habitual del significado y la concepción habitual [14].

La consideración de este punto llevó a los filósofos a una segunda etapa del debate sobre nuestra relación con teorías ancestrales y culturas exóticas. Se comprendía que si alguien era quineano convencido en su actitud hacia la noción de «significado» ni siquiera *desearía* preguntar «¿Querían decir lo mismo con '————'?» ¿Por qué se había armado, entonces, tanto alboroto? Probablemente por el deseo de saber cómo podrían asignar valores de verdad a las distintas manifestaciones. Queremos saber si debemos decir: «Aristóteles dijo principalmente falsedades sobre el movimiento» o decir más bien: «Aristóteles dijo principalmente cosas ciertas so-

[14] Putnam, *Mind, Language and Reality*, págs. 124-125.

bre lo que *él* llamaba 'movimiento', pero nosotros no creemos que exista tal cosa.» Además, queremos decir en algunos casos: «En esto Aristóteles cometió una pifia, incluso en sus propios términos», y en otros casos: «Aquí tenemos una afirmación que sería verdadera si en la física aristotélica hubiera *algo* verdadero, pero, por desgracia, se refiere a algo que no existe y, por tanto, es falso.» O, dicho de otra manera diferente, queremos distinguir entre las falsedades aristotélicas que son resultado de la inexistencia de aquello de que estaba hablando, y las que son resultado de la mala utilización de su propio aparato teórico —lo mismo que nosotros deseamos distinguir entre la falsedad de «Holmes vivía en Baker Street» y «Holmes estaba casado». Para establecer la distinción entre cuestiones internas (en que las respuestas se dan dentro de la cultura, la teoría, el relato, el juego) y externas (a las que se responde diciendo si es una cultura de la que somos miembros, una teoría que aceptamos, un relato que creemos, un juego que realizamos), podemos continuar con la idea de «referencia» y dejar de lado la de «significado». El «significado» sólo parecía importante porque ofrecía una forma de descubrir un objeto del mundo, del que luego podría determinarse si era el mismo o no era el mismo que algún objeto aprobado por nuestra propia cultura, teoría, relato o juego. Una vez que abandonamos la noción de significado, abandonamos también la idea de la referencia en cuanto determinada por el significado —de los «atributos definidores» de un término que hacen resaltar el referente del término.

La necesidad de descubrir objetos sin ayuda de definiciones, esencias y significados de los términos producía, pensaban los filósofos, una necesidad de una «teoría de la referencia» que no utilizara la maquinaria fregeana que Quine había hecho poner en duda. Esta exigencia de una teoría de la referencia llegó a asimilarse a la petición de una filosofía de la ciencia «realista» que restituyera la idea pre-kuhniana y pre-feyerabendiana de que la investigación científica progresaba averiguando cada vez más cosas sobre los mismos objetos [15]. No es que Kuhn ni Feyerabend hubieran negado esto, sino más bien que sus puntos de vista sobre la inconmensurabilidad de las teorías alternativas daba a entender que las únicas nociones de «verdad» y «referencia» que nosotros entendíamos realmente eran aquellas que se ponían en relación con un «esquema conceptual». Si no hubiera necesidad de que existiera un solo lenguaje de observación, común a todas las teorías alternativas, como habían sugerido

[15] Cfr. *ibíd.*, págs. 196 y ss.

Feyerabend y Kuhn (y Quine y Sellars habían contribuido a hacer creer, presentando nuevas razones), entonces había que abandonar la idea empirista de que siempre se podían dar definiciones operacionales de los términos teóricos. El ataque de Quine al primer dogma había dado carácter dudoso a la idea de «definición», y su ataque holista a la segunda, combinado con la afirmación de Sellars de que «lo dado al sentido» era cuestión de aculturación, hicieron a la idea de «definición operacional» doblemente dudosa. Las implicaciones antirreduccionistas de toda esta polémica antiempirista eran tales que empezó a resultar intelectualmente respetable algo que se parecía mucho al idealismo. Parecía posible decir que la cuestión de qué era real o verdadera no se iba a solucionar independientemente de un determinado marco de referencia conceptual y, a su vez, esto parecía sugerir que quizá no existiera realmente nada, aparte de esos marcos de referencia [16]. Por eso, la idea de que *podíamos* encontrar una matriz de investigación común —algo que abarcara todos los «marcos de referencia conceptuales» actuales y posibles— resultó muy atractiva a quienes tenían la sensación de que debía de haber *algo* verdadero en la desfasada idea prequineana y prekuhniana de que se encontraban representaciones de la naturaleza cada vez más precisa en su espejo.

No obstante, este clamor en torno al «idealismo» no es más que un pretexto para desviar la atención. Una cosa es decir (absurdamente) que construimos objetos utilizando palabras y otra muy distinta decir que no sabemos cómo encontrar una forma de describir una matriz duradera de la investigación pasada y futura de

[16] En realidad, esta idea sólo se les ocurrió a los *críticos* de Quine y Kuhn, que luego la tacharon de ser una *reductio*. Cfr. Suppe, «Search for Philosophical Understanding», pág. 151: «... si la ciencia considera siempre al mundo a través de una matriz disciplinar... ¿no está entonces Kuhn comprometido con cierta forma de idealismo antiempírico?» Israel Scheffler arremete contra el idealismo en el capítulo 1 de su *Science and Subjectivity* (Indianapolis, 1976), pero en los capítulos siguientes me parece que dejar ver cómo, aun después de que hayamos abandonado la idea de lo «dado» de C. I. Lewis y la mayor parte de la mitología empirista expuesta en el capítulo cuarto, no podemos tener control de la teoría mediante la observación (en la forma de control de las creencias menos controvertidas sobre las que lo son más). La afirmación de Scheffler (pág. 39) de que «tenemos simplemente una falsa dicotomía en la idea de que la observación ha de ser o una pura confrontación con algo dado e indiferenciado, o una cosa distinta tan contaminada conceptualmente como para hacer indirecta cualquier comprobación observacional de una hipótesis» parece precisamente lo que el mismo Kuhn desearía decir al respecto. Como me ha comentado Michael Williams, Kuhn no se interesa por el escepticismo, ni a favor ni en contra de él, y, sin embargo, los filósofos casi siempre le interpretan como si estuvieran presentando argumentos escépticos.

la naturaleza como no sea en nuestros propios términos —incurriendo así en una petición de principio contra «esquemas conceptuales alternativos». Casi nadie desea decir lo primero. Decir lo segundo, cuando se prescinde de la pavorosa retórica sobre «la pérdida de contacto con el mundo», es precisamente una forma de decir que nuestras concepciones actuales sobre la naturaleza son nuestra única orientación al hablar de la relación entre la naturaleza y nuestras palabras. Decir que tenemos que asignar referentes a los términos y valores de verdad a las oraciones a la luz de nuestras mejores nociones de lo que hay en el mundo, no pasa de ser un tópico. Decir que la verdad y la referencia son «relativas a un esquema conceptual» suena como si estuviéramos diciendo algo más que esto, pero no es así, con tal de que «nuestro esquema conceptual» se interprete simplemente como una referencia a lo que creemos ahora —la reunión de las concepciones que constituyen lo cultura de nuestros días. Esto es todo lo que nos permitiría entender por «esquema conceptual» cualquier argumento presentado por Quine, Sellars, Kuhn o Feyerabend. Sin embargo, Putnam, tras haber aclarado parte de la confusión creada por lo que Feyerabend decía del cambio de significado, pasó, por desgracia, a tratar la diferencia entre Feyerabend y él mismo no como una diferencia entre tomarse en serio la noción de «significado» y abandonarla, sino más bien como la diferencia entre una teoría del significado «idealista» y «realista».

Si se llega a ver cómo se desarrolló este pseudoproblema se comprenderá la etapa final del desarrollo de «la teoría de la referencia» a partir del denominado problema del cambio conceptual. Putnam dice, por ejemplo:

> El error de la teoría positivista de la ciencia es que está basada en una concepción del mundo que es idealista o tendente al idealismo, y que esa concepción no se corresponde a la realidad. Sin embargo, el elemento idealista del positivismo contemporáneo entra precisamente a través de la teoría del significado; por eso, dentro de cualquier crítica realista del positivismo se debe incluir al menos un esbozo de una teoría opuesta [17].

Con «una visión del mundo tendente al idealismo» Putnam quiere decir más o menos una concepción que «considere o tienda a considerar los 'hechos innegables' como meros hechos sobre *experiencias* actuales y potenciales, y todo lo demás como si fuera hablar en

---

[17] Putnam, *Mind, Language and Reality,* pág. 207.

forma muy derivada de las experiencias actuales y potenciales» (página 209). Por consiguiente, trata el deseo del positivismo de definiciones operacionales como si estuviera motivado no por la necesidad de garantizar una distinción analítico-sintética y, por tanto, un marco fijo para la investigación (tal como yo lo he hecho más arriba), sino por un deseo berkeleiano de evitar las preguntas sobre la relación entre experiencia y realidad independiente. En mi opinión, esto es una testarudez con relación a la historia, pero no tiene una importancia particular. Lo que es importante es la insistencia de Putnam en que «el positivista actual no tiene más derecho que Berkeley a aceptar la teoría y práctica científicas —es decir, su propio relato no lleva a ninguna razón para pensar que la teoría científica sea verdadera o que la práctica científica tienda a descubrir la verdad» (pág. 209). Putnam piensa que una de las características de los filósofos es que tienen «intuiciones o fuertemente antirrealistas o fuertemente realistas»:

> ... el antirrealista no ve nuestra teoría y la teoría de Arquímedes como dos descripciones aproximadamente correctas de algún ámbito fijo de entidades independientes de la teoría, y suele ser escéptico sobre la idea de «convergencia» en la ciencia —no piensa que nuestra teoría sea una descripción *mejor* de las *mismas* realidades que estaba describiendo Arquímedes. Pero si nuestra teoría es *solamente* nuestra teoría, entonces utilizar*la* para determinar si X está o no en la extensión de χρυσός es tan arbitrario como utilizar la teoría de Neanderthal para determinar si X está o no en la extensión de χρυσός. La única teoría que *no* resulta arbitrario utilizar es la que suscribe el propio hablante.
>
> El problema está en que para un antirrealista fuerte la *verdad* no tiene sentido como no sea en cuanto noción intrateórica... El antirrealista puede utilizar la verdad intrateóricamente en el sentido de una «teoría de la redundancia»; pero no puede disponer *extrateóricamente* de las nociones de verdad y referencia. La extensión de un término es precisamente aquello de lo que es verdad el término. En vez de intentar mantener la idea de extensión a través de un operacionalismo torpe, el antirrealista rechazaría la noción de extensión como rechaza la de verdad (en un sentido extrateórico). Como Dewey, por ejemplo, puede volver a una idea de «afirmabilidad garantizada» en vez de verdad... Luego puede decir que «X es oro (χρυσός)» se podía afirmar justificadamente en tiempo de Arquímedes y no se puede afirmar justificadamente hoy... pero la afirmación de que X estaba dentro de la extensión de χρυσός será rechazada por ca-

> rente de significado, como la afirmación de que «X es oro (χρυσός)» era verdadera (pág. 236).

Es difícil encontrar un filósofo que cumpla los criterios establecidos por Putnam para ser «antirrealista». Kuhn apunta en algunas ocasiones que le parece inaceptable la idea de «una descripción mejor de las mismas realidades», pero la mayoría de los filósofos que se han dejado influir por sus críticas a la filosofía positivista de la ciencia no han considerado necesario llegar tan lejos. Una vez más, los filósofos que tienen dudas sobre la noción de una teoría de las conexiones entre las palabras y el mundo, y que, como Sellars, se sienten inclinados a poner «verdadero» en relación con los marcos de referencia conceptuales, nunca dirían que carecen de sentido las afirmaciones que en opinión de Putnam deberían considerar como no significativas. Sellars interpretaría «que se puede afirmar justificadamente en nuestro marco de referencia conceptual pero que no es verdadero» como una referencia implícita a otro marco de referencia conceptual, quizá no inventado todavía, en el que la afirmación en cuestión no se pudiera afirmar justificadamente. Los filósofos que, como James y Dewey y Strawson, tienen dudas sobre «la teoría de la verdad de la correspondencia», no sienten ninguna simpatía por la noción de la naturaleza como maleable al pensamiento, ni por la inferencia de «no se puede ofrecer una descripción de una cosa que sea independiente de la teoría» a «no hay cosas que sean independientes de la teoría». Como suele ocurrir en los debates entre «idealistas» y «realistas», ambos bandos quieren que sea el contrario el encargado de presentar pruebas. El idealista afirma ser capaz de dar sentido satisfactorio a todo lo que quiera decir el sentido común, e incluso la filosofía del lenguaje, y se pregunta qué es lo que podrían añadir los realistas. El realista insiste en que la concepción idealista tiene consecuencias contraintuitivas de las que sólo nos puede librar una teoría de una relación llamada «correspondencia entre pensamiento (o palabras) y el mundo».

Putnam sigue tres líneas argumentales, cada una de ellas orientada a demostrar que hay una diferencia significativa entre el realista y su contrario y que la razón está del lado del realista. La primera es un argumento contra la interpretación de «verdadero» como si significara «que se puede afirmar justificadamente» o cualquier otra idea «suave» que tenga que ver con las relaciones de justificación. Se supone que esto demuestra que sólo una teoría sobre las relaciones entre las palabras y el mundo puede ofrecer una interpretación satisfactoria. La segunda es un argumento en el

sentido de que aquel tipo de hecho sociológico que necesite explicación —la fiabilidad de los métodos clásicos de la investigación científica, o la utilidad de nuestro lenguaje como instrumento para valernos con el mundo— sólo se puede explicar con razones realistas. La tercera es un argumento de que sólo el realista puede evitar pasar de «muchos términos utilizados en las ciencias del pasado no hacían referencia» a «es más que probable que ninguno de los términos utilizados por nuestros científicos haga referencia», conclusión que le parece inadmisible[18]. Creo que sólo el tercero de estos argumentos se enfrenta con un enemigo real. Para examinar este argumento, estudiaré, en las tres siguientes secciones, la referencia, la verdad y el relativismo. Pero, antes de acabar la presente sección, trataré de formular una crítica a los dos primeros argumentos de Putnam que sirva para hacer ver por qué sólo merece tomarse en serio el tercero.

La afirmación de Putnam de que ninguna idea del tipo «que se puede afirmar justificadamente» tendrá los mismos rasgos sintácticos que «verdadero» está perfectamente justificada. Pero no está claro que tenga que ver con ninguna afirmación hecha por algún filósofo. Los filósofos que sugieren que «verdadero» quiere decir «que se puede afirmar justificadamente» generalmente o (a) hacen la noción de verdad relativa a un lenguaje, teoría, etapa de investigación o esquema conceptual, o (b) explican que no necesitamos una noción de «verdadero» en el momento en que tenemos la de «que se puede afirmar justificadamente». En otras palabras, o recomiendan una revisión de nuestro uso normal de «verdadero» o aconsejan que abandonemos por completo el término. Como señala el propio Putnam, los argumentos contra nociones como las de *análisis* de «verdadero» son tan fáciles como los argumentos de Moore contra los intentos de definir «bueno», y por razones muy semejantes. «Verdadero pero que no se puede afirmar justificadamente» tiene tanto sentido como, por ejemplo, «bueno pero no conducente a la máxima felicidad» o «bueno pero condenado por todas las culturas existentes hasta ahora». Los filósofos que tienen interés (cosa que *no* tienen Tarski ni Davidson) en decirnos sobre la verdad algo que explique o garantice el éxito de nuestra *búsqueda* de la verdad son como los filósofos que quieren decirnos sobre «bueno» algo más que se utiliza para recomendar, algo que expli-

[18] Putnam, «Waht Is 'Realism'?», *Proceedings for the Aristotelian Society,* página 194. (Reimpreso con algunos cambios en su obra *Meaning and the Moral Sciences* [Londres, 1978], libro que apareció cuando este capítulo ya estaba escrito.)

que o garantice el progreso moral. Pero quizá no se pueda decir mucho en este sentido. Utilizando, una vez más, la analogía con la filosofía moral, la comprensión de por qué «bueno» es indefinible o de cómo se utiliza no sirve en absoluto para explicar que una acción buena es aquella que se corresponde con la Forma del Bien, o con la Ley Moral. Tampoco nos sirve de nada que nos digan que las afirmaciones verdaderas se corresponden con la forma de ser del mundo [19].

La postura adecuada del «antirrealista» es admitir simplemente que nada va a explicar la «verdad independiente de la teoría», igual que nada podrá explicar «bondad no instrumental» o «belleza no-funcional», y hacer recaer sobre Putnam la obligación de presentar las pruebas. La reacción podría consistir, más o menos, en preguntar: «¿Qué perderíamos si no tuviéramos una noción de verdad ahistórica, independiente de la teoría?» Esta pregunta parece tan razonable como la siguiente: «¿Qué conseguimos cuando Sócrates nos enseñó a usar 'bueno' de tal manera que tuvieran sentido 'malo para mí, pero sin embargo bueno', 'malo para Atenas, pero sin embargo bueno' e incluso 'detestable para los dioses, pero sin embargo bueno'?» Es incuestionable que tenemos estas nociones de verdad y bondad —nociones que se mantienen independientes de todas las preguntas sobre la justificación. Es igualmente incuestionable que esta noción de verdad tiene ciertas propiedades que no va a tener ninguna noción de afirmabilidad o justificación (por ejemplo, por citar algunas de las que menciona Putnam: si una afirmación es verdadera, también lo son sus consecuencias lógicas; si dos afirmaciones son verdaderas, también lo es su unión; si una afirmación es verdadera ahora, lo es siempre). Es posible que esta idea no sea mucho más antigua que Sócrates y Aristóteles, antes de los cuales la «consecuencia lógica» habría resultado ininteligible. Pero, cualquiera que sea su origen, el hecho de que tengamos esta idea no es de por sí garantía de que haya una teoría filosófica sobre ella. La mayor parte de lo que, en los libros de filosofía, pasa por estudio de la «verdad» es, de hecho, estudio de la justificación, lo mismo que la mayoría de lo que pasa por estudio de la «bondad» es estudio del placer y el dolor. El precio a pagar por

---

[19] Véase Nelson Goodman, «The Way the World Is», en su *Problemas and Projects* (Indianapolis, 1972), págs. 24-32, especialmente pág. 31: «El mundo es de muchas maneras, y cada descripción verdadera capta una de ellas.» Creo que la afirmación de Goodman tiene su mejor explicación en Davidson, «On the Very Idea of a Conceptual Scheme» (estudiado más adelante) y «Mental Events» (mencionado en el capítulo cuarto, sección, 5 anteriormente).

distinguir netamente los *transcentalia* de sus réplicas de sentido común puede ser que no deje material para la construcción de teorías, ni problemas que resolver.

Sin embargo, Putnam señala muy explícitamente que hay problemas que resolver que no tienen nada que ver con la justificación, y el segundo de los argumentos que he mencionado es una respuesta directa a la pregunta «¿*Para* qué necesitamos la idea de verdad, en cuanto opuesta a la de justificación?» La respuesta de Putnam es que la necesitamos para explicar la fiabilidad de nuestros procedimientos de investigación. Más en concreto, la necesitamos para explicar el hecho de la «convergencia» en la ciencia —el hecho de que las antiguas teorías falsas presenten, según se van aproximando a nuestro propio tiempo, aproximaciones cada vez más acertadas a nuestras teorías actuales [30]. La objeción obvia a este punto de partida es que esta «convergencia» es un artificio inevitable de la historiografía. Parece claro que habrá siempre una forma natural de contar la historia de la sucesión de teorías (o de la sucesión de religiones o formas de gobierno) que haga ver cómo nuestros predecesores fueron avanzando gradualmente, aunque a trancas y barrancas, hacia donde nos encontramos ahora. No hay ninguna razón para pensar que el antirrealista no vaya a tener algo que decir sobre los efectos causales sobre nuestros predecesores de los objetos de que habla nuestra actual teoría. Puede describir también cómo estos objetos contribuyeron a producir descripciones justificadas pero falsas de sí mismos, a las que seguirían descripciones igualmente justificadas, incompatibles y ligeramente mejores, y así sucesivamente hasta nuestros días. Sin embargo, si la «convergencia» se interpreta no como un hecho de la historia de las diversas disciplinas sino como algo relacionado con los resultados de las comprobaciones de las nuevas teorías, entonces quizá parezca que el realista está en mejor situación para hacer uso de ello. Boyd, que ha colaborado con Putnam en la creación de esta línea de argumentación, interpreta «convergencia» como la «fiabilidad» de un principio como éste:

> habría que indagar, a la luz del conocimiento teórico disponible, en qué circunstancias las afirmaciones causales hechas por la teoría podrían resultar erróneas, bien porque otros mecanismos

---

[30] Véase Putnam, «Reference and Understanding», en Meaning and the *Moral Sciences*, págs. 97-119. He tenido conocimiento de este trabajo gracias a unas pruebas de imprenta que me proporcionó amablemente el profesor Putnam.

> causales plausibles a la luz de los conocimientos existentes pudieran estar operando en vez de los indicados por la teoría, o bien porque podría esperarse plausiblemente que los mecanismos causales ya conocidos entraran en conflicto con los requeridos por la teoría en formas no previstas por la teoría [21].

Es difícil imaginar que alguien presente objeciones a este principio, por lo que la discusión no comienza hasta que Boyd afirma que sólo podemos explicar que este principio lleve a resultados prácticos «basándose en una comprensión realista de las teorías colaterales relevantes»:

> Supongamos que alguien «adivina» siempre dónde es más probable que las teorías se equivoquen experimentalmente preguntando dónde es más probable que sean falsas en cuanto explicaciones de las relaciones causales, dada la suposición de que las leyes aceptadas en la actualidad representan un conocimiento causal probable. Y supongamos que su procedimiento de adivinación resulta —que las teorías tienen realmente muchas probabilidades de equivocarse— de dar lugar a predicciones experimentales falsas —precisamente donde un realista esperaría que así ocurriera...—. ¿Qué explicación es posible, además del realismo científico? Ciertamente, *no* el mero resultado de tradiciones científicas convencional o arbitrariamente adoptadas... A no ser que, cosa que no sugeriría ningún empirista, el mundo sea moldeado por nuestras *convenciones,* es imposible que la fiabilidad de este principio pueda ser una cuestión de convención, meramente (pág. 12).

Aquí Boyd confunde el sentido en que un procedimiento es fiable con relación a una prueba independiente (como los termómetros son indicadores fiables del frío que hace fuera) con el sentido en que un procedimiento es fiable porque no podemos imaginar una alternativa. Comprobar las teorías nuevas por medio de las antiguas no es un procedimiento opcional. ¿De qué otra manera podríamos comprobarlas? El hecho de que las teorías nuevas se equivoquen muchas veces precisamente donde las anteriores decían que podía ocurrir tal cosa no es algo que necesite explicación. Necesitaría explicación si se equivocaran en alguna otra parte. Sólo podríamos hacer que pareciera que el «no-realista» es incapaz de explicar la relevancia de las teorías antiguas para comprobar las nuevas si aceptáramos que este hombre de paja afirmaba que las nuevas teorías

[21] Richard N. Boyd, «Realism, Underdetermination and a Causal Theory of Evidence», *Nous,* 7 (1973), 11.

llegan completamente equipadas, con lenguajes observacionales igualmente nuevos, procedimientos de comprobación y principios reguladores propios. Pero una «nueva teoría» no es más que un pequeño cambio en una amplia red de creencias. Su verdad, como decía James, es en gran parte cuestión de su capacidad de realizar «una función de matrimonio» entre el depósito de la antigua verdad y la «anomalía» que la sugirió en primer lugar. Sólo algunos de los conceptos más forzados de Feyerabend nos indican que debamos conceder a una nueva teoría inmunidad ante cualquier test basado en los resultados de otra antigua. Estas sugerencias no tienen nada que ver con la cuestión de si necesitamos una idea de «verdad» además de la de «afirmabilidad justificada».

4. La referencia

Para demostrar que *sí* necesitamos esta idea hemos de dirigir la atención al tercer argumento de Putnam, que tiene la virtud de presentarnos un «antirrealista» más plausible que los que hemos considerado hasta ahora. Este argumento gira en torno a lo que Putnam llama la necesidad de bloquear «la desastrosa meta-inducción que concluye 'ningún término teórico hace referencia nunca'» [22].

> ¿Qué ocurre si aceptamos una teoría desde cuyo punto de vista los electrones son como el *flogisto?*
>
> Entonces tendremos que decir que los electrones no existen realmente. ¿Qué pasa si esto vuelve a ocurrir? ¿Y si todas las realidades teóricas postuladas por una generación (moléculas, genes, etc., así como los electrones) «no existen», invariablemente, desde el punto de vista de la ciencia más reciente? Se trata, por supuesto, de una forma del antiguo «argumento a partir del error» de los escépticos —¿cómo sabemos que no estamos en el error *ahora?* Pero es la forma en que el argumento del error constituye una preocupación *seria* para muchas personas de hoy día, y no solamente una «duda filosófica».
>
> Una razón por la que se trata de una preocupación seria es que en algún momento resulte abrumadoramente irresistible la siguiente meta-inducción: *igual que ningún término de la ciencia de hace más de cincuenta años* (o los que sean) *hacía referencia, también resultará que ningún término utilizado ahora* (con excepción, quizá, de los términos de observación, si es que los hay) *hace referencia* (págs. 183-184).

[22] Putnam, «Waht Is Realism?», pág. 194.

Putnam dice que cerrar el paso a esta meta-inducción constituye «evidentemente un desideratum para la Teoría de la Referencia». Esto resulta enigmático, por dos razones. En primer lugar, no está claro qué punto de vista filosófico podría demostrar que había acabado el cambio revolucionario dentro de la ciencia —que *no* seremos para nuestros descendientes como nuestros primitivos antepasados animistas son para nosotros. ¿Cómo podemos alejarnos de nuestra propia cultura y evaluar su lugar en relación con el fin de la investigación? En segundo lugar, aun cuando hubiera un punto de vista filosófico semejante, no es claro cómo podría proporcionarlo la teoría de la referencia. Supongamos que tenemos intuiciones decisivas sobre la respuesta a preguntas como «Si Jones no hizo las grandes hazañas que se le atribuyen, pero resulta que Smith (anteriormente desconocido para la historia) hizo la mayoría de ellas, ¿'Jones' se refiere a Jones o a Smith?» Supongamos, menos plausiblemente, que nuestras intuiciones sobre estos casos enigmáticos son lo suficientemente firmes como para llevarnos a una teoría general que nos informe, sin recurrir a la intuición, de que Dalton y nosotros nos referimos a las mismas cosas con «molécula» mientras que «fluido calórico» no se utilizó nunca para referirse al movimiento de las moléculas [28]. Dicha teoría podría hacer retroceder a las premisas de la meta-inducción sugerida todavía más hacia el pasado —pues ahora quizá tuviéramos que remontarnos cinco mil años en vez de sólo cincuenta para buscar una ciencia ninguno de cuyos términos teóricos hiciera referencia. Pero esto difícilmente satisfaría a un escéptico que hubiera llegado a preocuparse seriamente por la situación de nuestra ciencia con relación, por ejemplo, a la civilización galáctica del futuro. Es de suponer que este escéptico sólo se sintiera satisfecho con una teoría de la referencia que demostrara que en todas las etapas y lugares los científicos se habían venido refiriendo en su mayor parte a las mismas cosas, privando así a la meta-inducción en cuestión de *cualquier* premisa interesante.

Ahora bien, en un sentido obvio, sabemos perfectamente —antes de ninguna teoría— que se han estado refiriendo a las mismas cosas. Todos ellos intentaban vérselas con el mismo universo, y hacían referencia a él, aunque, indudablemente, muchas veces bajo descripciones absurdas e improductivas. El descubrir, como consecuencia de la próxima revolución científica, que no hay genes, moléculas, electrones, etc., sino únicamente sacudidas espacio-tempora-

[28] He tomado el ejemplo de Rom Harré, *The Principles of Scientific Thinking* (Chicago, 1970), pág. 55.

les, o sugestiones hipnóticas producidas por los hipnotizadores galácticos que han manipulado a nuestros científicos desde la época de Galileo, o cualquier otra cosa, no haría, sin embargo, que perdiéramos el contacto con el mundo o con nuestros antepasados. Nos pondríamos a contar la misma historia de la aparición de descripciones mejores del mundo a partir de descripciones falsas, confusas, inútiles del mundo que contamos, por ejemplo, sobre el nacimiento de la ciencia milesia —en ambos casos, sería la historia de un triunfo de la razón, y en ambos casos sería la aplicación de la razón al mismo mundo. En general, no hay ninguna sugerencia que pudiéramos hacer sobre cómo resultará ser el mundo, que parezca servir de base para una «preocupación seria». (La sugerencia de que pudiera haber alguna descripción ininteligible del mundo —descripción que estuviera formulada dentro de un esquema conceptual intraducible— es una preocupación completamente distinta, que no tiene ninguna necesidad de apoyo inductivo, ni ninguna posibilidad de conseguirlo. Me ocupo de esta otra preocupación al hablar del ataque de Davidson a la idea de «esquema conceptual», en la siguiente sección.)

Sin embargo, parece que esta idea de «mala descripción de la misma cosa» elude un problema. La «referencia» parece una cuestión de o-todo-o-nada. Puede parecer mejor decir que el «fluido calórico» o se refiere o no se refiere al movimiento de las moléculas que decir que «fluido calórico» y «movimiento de las moléculas» son dos abreviaturas de descripciones (mejor y peor) del mismo fenómeno, sea lo que sea lo que resulte ser. Podemos tener la impresión de que hacer referencia al movimiento de las moléculas debería ser como hacer referencia a personas individuales o a objetos físicos de tamaño medio —o los distinguimos o no—. Esta impresión se ve fortalecida si hemos estado leyendo últimamente a Tarski, y todavía más si no nos sentimos inclinados a considerar las afirmaciones teóricas en la forma cuasi-instrumentalista de Sellars —como principios materiales de inferencia que hacen posible enunciar otras afirmaciones. Desearemos entonces que oraciones como «El calor es el movimiento de las moléculas» sea verdad en la «correspondiente» forma clara en que lo es «El blanco es el color característico de la nieve». Por eso puede parecer importante, en orden a la comprensión de «cómo funciona el lenguaje», pensar en términos de expresiones «que distinguen realidades» en vez de que sean simplemente utilizadas «para describir la realidad». Supongamos que es así. Sin embargo, cuesta ver por qué tenemos que hacer más para calmar al escéptico que para ser «Whiggish» en nuestra historio-

grafía. Podemos escribir las cosas de tal manera que hagamos que hasta el más primitivo de los animistas hable, por ejemplo, del movimiento de las moléculas, del radio, de los genes, o lo que sea. Con eso *no* calmamos su temor de que no existan las moléculas, pero tampoco se conseguirá eso con *ningún* descubrimiento sobre la forma en que se relacionan las palabras con el mundo, pues «el mundo», tal como es conocido por esa teoría, es precisamente tal como es conocido por la ciencia de la época.

Hasta ahora hemos estado ante el siguiente dilema: o se recurre a la teoría de la referencia para subrayar el éxito de la ciencia contemporánea, o de lo contrario es solamente cuestión de decisión sobre cómo escribir la ciencia (en vez de presentar un «fundamento filosófico» para dicha historiografía). La primera tarea parece excesiva, y la segunda demasiado ligera como para merecer el título de «teoría». En este punto, sería conveniente preguntar de dónde se originó por primera vez la idea de «teoría de la referencia». ¿Qué cosa es esa sobre la que habría que hacer recaer la tarea de impedir el escepticismo, haciendo así el trabajo que siempre habíamos esperado que realizara la epistemología? En mi opinión, la idea de que existe tal teoría procede de la unión de dos consideraciones muy diferentes, a saber:

> el hecho, señalado por Kripke, Donnellan, y otros, de que hay contraejemplos del criterio de referencia de Searle-Strawson —es decir, que S, en su uso de «X», hace referencia a cualquier entidad que sacara el máximo provecho de sus creencias centrales sobre la verdad X[24].
>
> El hecho de que la suposición habitual (Frege, Searle, Strawson) de que el significado, en el sentido de creencias o intenciones (o, en sentido más general, entidades en las cabezas de los usuarios de palabras), determina la referencia, permite pensar que cuantas más falsas creencias tengamos menos «en contacto con el mundo» estamos.

Las dos consideraciones unidas sugieren que la idea «intencionalista» habitual sobre cómo las palabras están conectadas con el mundo al

[24] El estudio más reciente de estos contra-ejemplos es el de Keith Donnelan, «Proper Names and Identifying Descriptions» y Saul Kripke, «Naming and Necessity», ambos aparecidos en *Semantics of Natural Language,* ed. Davidson and Harman (Dordrecht, 1972). Véase Stephen P. Schwartz, ed., *Naming, Necessity and Natural Kinds* (Ithaca, 1976), donde hay nuevos artículos que intentan construir una teoría general que permita explicar estos contra-ejemplos, y una interesante exposición de las obras publicadas, en la «Introducción» del editor.

mismo tiempo que es errónea en casos particulares resulta filosóficamente desastrosa. Así, en la filosofía «impura» del lenguaje, casi ha llegado a ser un dogma que las doctrinas de aire «idealista» de Quine, Wittgenstein, Sellars, Kuhn, Feyerabend (y otros héroes de este libro) se deben refutar retrocediendo a los primeros principios de la semántica, derribando la teoría «intencionalista» de la referencia de Frege, y sustituyéndola por otra mejor [25]. La idea es que si el mundo extiende la mano y se traba con el lenguaje en relaciones factuales (por ejemplo, causales), siempre estaremos «en contacto con el mundo», mientras que en la concepción fregeana estamos en peligro de perder el mundo, o quizá no hayamos estado nunca trabados con él.

Sin embargo, deberíamos sospechar de la noción de un enfrentamiento entre una teoría de la referencia antigua —intencionalista— y otra nueva, «causal» (o, más generalmente, no-intencionalista y, por consiguiente, «realista»). El enfrentamiento se produce por la equivocidad de «referencia». El término puede significar o *a)* una relación factual que se da entre una expresión y alguna otra parte de la realidad, tanto si alguien sabe que se da como si no, o *b)* una relación puramente «intencional» que pueda darse entre una expresión y un objeto inexistente. Llamaremos a lo primero «referencia» y a lo segundo «hablar de». No podemos hacer *referencia* a Sherlock Holmes, pero podemos *hablar de* él, y lo mismo ocurre con el flogisto. «Hablar de» es una idea de sentido común; «referencia» es un término de la ciencia filosófica. «Hablar de» abarca ficciones y realidades, y es inútil a efectos realistas. La suposición de que las creencias de las personas determinan qué es aquello de lo que están hablando funciona tan bien o tan mal en el caso de las cosas que existen como en el de las que no existen, con tal de que no se planteen preguntas sobre qué es lo que existe. En una comunidad donde no haya teorías encontradas (físicas, históricas, «ontológicas», o cualesquiera otra), pero en la que sea bien conocido que algunas de las personas y cosas de las que se habla existen, mientras que otras son ficciones, podríamos utilizar ciertamente los criterios de Searle-Strawson. *Estamos* hablando sobre aquello, sea lo que sea, de lo que la mayoría de nuestras creencias son verdaderas. Los casos enigmáticos (casos en los que la intuición nos dice que las personas *no* están hablando de la entidad, cualquiera que sea, que hace verdaderas la mayoría de sus creencias) sólo se presentan cuando sa-

[25] Putnam, «The Meaning of 'Meaning'» y su «The Refutation of Conventionalism» (ambos incluidos en su *Mind, Language and Reality*) son las expresiones más completas de esta actitud.

bemos algo que ellos no saben. Así pues, si descubrimos que hubo un hombre desconocido hasta ahora, llamado Smith, que realizó el 99 por 100 de las hazañas atribuidas al mítico Jones, pero que las historias sobre Jones se formaron realmente en torno a un hombre llamado Robinson, quizá queramos decir que cuando hablamos de Jones estamos hablando *realmente* de Robinson y no de Smith.

Si se confunde esta idea de «hablar realmente de» con la de *referencia,* resulta fácil pensar (como hacen Putnam y Kripke) que tenemos «intuiciones» sobre la referencia, intuiciones que podrían ser la base de una «teoría de la referencia» no intencional e «independiente de la teoría». Pero esto es pensar que la prgunta

a) ¿Cuál es la mejor forma de expresar la falsedad de las creencias habituales sobre Jones —diciendo que no trataban de nada en absoluto, que eran verdades sobre una ficción, o falsedades sobre una realidad?

debería contestarse partiendo de una respuesta a

b) ¿Hay una entidad del mundo que esté conectada con nuestro uso de «Jones» por la relación de «referencia»?

Según el punto de vista que adoptan Putnam y Kripke, *b)* es una pregunta razonable, y es antecedente de *a).* Según ello, el responder a *a)* no es cuestión de conveniencia historiográfica o de exposición, sino de hecho —un hecho determinado por la respuesta a *b).* Según el punto de vista que yo considero más recomendable, no se plantea *b).* El único problema factual en todo esto se refiere a la existencia o no existencia de varias entidades de las que se habla. Una vez que hayamos tomado una decisión sobre la última pregunta factual, podemos adoptar una de estas cuatro actitudes hacia las creencias en que (según el criterio de sentido común de Searle-Strawson) la persona en cuestión está hablando de realidades inexistentes. Podemos

1. Declararlas a todas falsas (Russell) o sin valor de verdad (Strawson).
2. Dividirlas entre las que son falsas o sin valor de verdad porque no tratan de nada y las que «tratan realmente» de algo real y que pueden ser verdaderas

o

3. Separarlas en las que son falsas o sin valor de verdad porque no tratan de nada y las que «tratan realmente» de rea-

lidades de ficción y, que por tanto, pueden ser quizá verdaderas

o

4. Combinar las estrategias (2) y (3).

El «realmente» y «tratar realmente de» refleja nuestro alejamiento del criterio de Searle-Strawson, pero no constituye una invocación a nuestras intuiciones en relación con una cuestión de hecho. Es como la idea de «ha estado realmente bien hacer tal cosa» utilizada cuando, a pesar de que alguien ha actuado aparentemente en forma condenable, una opinión más amplia e informada indica que convendría dejar de lado los criterios morales de sentido común. En el caso moral, no estamos teniendo una intuición sobre la conexión factual entre la acción y la Forma del Bien; simplemente, estamos reformulando nuestra descripción de la situación de una manera que evite la paradoja y consiga el máximo de coherencia. De la misma manera en el caso de las decisiones sobre quién era realmente el que estaba hablando y sobre qué hablaba.

Puede parecer que la diferencia entre la concepción que estoy defendiendo y la de Putnam-Kripke se basa en la cuestión del «Meinongianismo» —de si es posible referirse a ficciones. En el sentido en que la utilización del término *referirse* está dirigida por la inferencia de

«N» hace referencia y N es ϕ

a

N existe (no es una ficción),

no podemos referirnos a ficciones. Ésta es la forma habitual de utilizar el término *referirse,* y no tengo intención de usarlo de forma diferente. Pero la conclusión que extraigo del hecho de que esta condición define la noción de «referencia» es que «referencia» no tiene nada que ver en concreto ni con «hablar de» ni con «hablar realmente de». La «referencia» sólo se da cuando alguien ha tomado una decisión sobre las diversas estrategias utilizadas para expresar el error que encuentra en el mundo —la decisión entre (1)-(4), anteriormente— y luego desea dar al resultado de su decisión forma «canónica», es decir, formularlo en un lenguaje que utilice como

matriz la lógica cuantificacional estándard. Esto es lo que quería decir cuando decía que «referencia» es un término propio de una ciencia. Es también la razón por la que es algo de lo que no tenemos intuiciones. Por eso, llego a la conclusión de que la «intuición» con que se enfrenta el criterio de Searle-Strawson es meramente la intuición de que cuando hay discusión sobre lo que existe puede haber discusión sobre qué es «que se hable realmente de algo», y que el criterio para «realmente de» no es el de Searle-Strawson.

¿Cuál es, entonces? No hay respuesta a esta pregunta, no existe tal «criterio». Las consideraciones que dictan la elección entre las estrategias (1)-(4) de más arriba son tan diversas que no tiene sentido buscar un criterio. Podríamos tener la tentación de decir que «hablar realmente de» es una relación que se puede dar entre una expresión y lo que *nosotros* pensamos que existe, en oposición a «hablar de», que se da entre una expresión y lo que su usuario cree que existe, y a «se refiere», que sólo puede darse entre una expresión y lo que existe real y verdaderamente. Pero esto sería un error, pues, una vez más, no sólo podemos hablar de realidades inexistentes, sino que es posible que se descubra que hemos estado hablando *realmente* de realidades inexistentes. Hablar realmente de X no es lo mismo que hablar de un X real. En este caso «realmente» sólo intenta «colocar» la ignorancia relativa de la persona en cuestión en el contexto del conocimiento relativamente mayor proclamado por el hablante. Son tantas las formas de hacer esto como los contextos del discurso. Por ejemplo: «Crees que estás hablando realmente de Thales, pero estás hablando realmente de un cuento contado por Herodoto»; «Crees que estás hablando de tu analista, pero estás hablando realmente de ti mismo»; «Crees que estás hablando de una deidad de ficción llamada Artemis, pero estás hablando realmente de una mujer de carne y hueso que vivió en Tebas en el siglo IX antes de Cristo»; «Crees que estás hablando del litio, pero estás hablando realmente de la criptonita»[26].

Por eso, pienso que la búsqueda de una teoría de la referencia representa una confusión entre la búsqueda «semántica» desesperada de una teoría general de qué es aquello de lo que «están hablando realmente» las personas, y la búsqueda «epistemológica» igualmente desesperada de una forma de refutar al escéptico y confirmar nuestra afirmación de que estamos hablando de realidades no ficticias.

[26] He desarrollado este punto de vista sobre la referencia con mayor amplitud en «Realism and Reference», *The Monist,* 59 (1976), págs. 321-340. Véase una crítica paralela a Putnam en este punto en Arthur Fine, «How to Compare Theories: Reference and Change», *Nous,* 9 (1975), págs. 17-32.

No es necesario atender ninguna de estas dos peticiones para una filosofía «pura» del lenguaje de estilo davidsoniano. La primera consiste, más o menos, en exigir un procedimiento de decisión para resolver los casos difíciles de la historiografía, descripción antropológica y semejantes —casos donde sólo servirán el tacto y la imaginación. La segunda es la exigencia de algún punto de vista transcendental ajeno a nuestro actual conjunto de representaciones, desde el cual podamos inspeccionar las relaciones entre esas representaciones y su objeto. (Ésta era la exigencia que Berkeley decía que no podríamos satisfacer, que Kant satisfizo únicamente llamando al mundo «apariencia», y que la imagen del Espejo de la Naturaleza nos hace pensar que deberíamos ser capaces de satisfacer.) La pregunta «¿Qué determina la referencia?» es ambigua, pudiendo ser una pregunta sobre el procedimiento más indicado para comparar amplios conjuntos coherentes de creencias falsas (otras épocas, culturas, etc.) con los nuestros y una pregunta sobre cómo refutar al escéptico. Los debates sobre las teorías de la referencia reciben su concreción de los intentos de responder a la primera parte de la pregunta, y su interés filosófico de las indicaciones de que podría responder de alguna manera a la segunda. Pero no hay nada que pueda refutar al escéptico —nada puede hacer lo que esperaba hacer la epistemología. Sólo descubrimos cómo funciona el lenguaje dentro de la teoría actual del resto del mundo, y no se puede utilizar una parte de la propia teoría actual para confirmar el resto de ella. La teoría de la referencia es tan desesperada como una «teoría de la constitución transcendental del objeto» a estos fines.

Putnam (en una conferencia pronunciada cuando este capítulo estaba ya prácticamente terminado) se ha retractado en gran parte de su «realismo metafísico» —el proyecto de explicar la referencia que consigue su objetivo recurriendo a algún medio que no presuponga ese logro. En esta conferencia, dice lo mismo que yo vengo señalando sobre las «teorías causales». Afirma que lo que quería el realista metafísico, pero sin poder conseguirlo, era una concepción de la «verdad como radicalmente no espistémica» —es decir, «una concepción en la que la teoría que es 'ideal' desde el punto de vista de su utilidad operacional, elegancia y belleza internas, 'plausibilidad', simplicidad, 'conservadurismo', etc., *podría ser falsa*»[27]. El realista metafísico *pensaba* que tenía que decir esto porque parecía ser la

[27] Putnam, «Realism and Reason», *Proceedings of the American Philosophical Association,* vol. 50 (1977), pág. 485. (Este artículo ha vuelto a aparecer en *Meaning and the Moral Sciences.* El pasaje en cuestión esta en la página 125.)

única forma de separar claramente «verdadero» de «que se puede afirmar justificadamente». Pero, como dice Putnam, aun cuando definamos «verdadero» *à la* Tarski en términos de la relación de satisfacción, seremos capaces de reproducir *cualquier* conjunto de creencias sobre el mundo en términos de esta relación. Además, habrá muchas formas *diferentes* de poder hacer esto, y *no hay constricciones sobre las formas de hacerlo como no sean las constricciones sobre las teorías en general.* Nuestra mejor teoría sobre aquello a que nos estamos refiriendo es mero resultado de nuestra mejor teoría sobre las cosas en general. Como dice Putnam: «... un teoría 'causal' de la referencia no sirve (no serviría) para nada aquí: en la imagen realista metafísica, el cómo pueden hacer referencias las 'causas' es tan problemático como la forma en que pueda hacerlo 'gato'» [28]. De la misma manera, independientemente de cuál sea la relación no-intencional que sustituya a «causa» es nuestra explicación de cómo las cosas del contenido llegan a determinar la referencia de las representaciones que constituyen el esquema, nuestra teoría sobre qué es aquello de lo que está hecho el mundo produciría, superficialmente, una teoría autojustificativa sobre esa relación.

5. Verdad sin espejos

La retractación de Putnam equivale a decir que no hay manera de conseguir que una disciplina empírica haga lo que no pudo hacer la filosofía transcendental —o sea, decir algo sobre el esquema de representaciones que estamos empleando que aclare su conexión con el contenido que deseamos representar. Pero si esto no es posible, podemos aceptar la afirmación de Davidson de que debemos abandonar por completo la distinción esquema-contenido. Podemos admitir que no es posible conseguir que la noción de «esquema» haga lo que los filósofos han querido hacer tradicionalmente con ella —es decir, aclarar ciertas constricciones especiales que ejerce la «racionalidad» y que explican por qué nuestras teorías ideales han de «corresponder a la realidad». Putnam está de acuerdo ahora con Goodman y Wittgenstein: pensar en el lenguaje como una imagen del mundo —un conjunto de representaciones que la filosofía ha de presentar como si estuvieran en una especie de relación no intencional con lo que representan— *no* sirve para explicar cómo se aprende o se entiende el lenguaje. Pero, al menos en sus escritos

[28] Putnam, «Realism and Reason», pág. 486 (págs. 126 en *Meaning and the Moral Sciences).*

anteriores a esta retractación, pensaba que podríamos utilizar esta imagen del lenguaje a efectos de una epistemología naturalizada; el lenguaje-como-imagen no era un concepto útil para entender cómo se utilizaba el lenguaje, pero servía para explicar el éxito de la investigación, lo mismo que «un mapa cumple sus objetivos si se corresponde de alguna forma adecuada con una parte concreta de la tierra». Putnam está realizando así el mismo movimiento realizado por Sellars y Rosenberg. Éstos identifican «verdadero» con «que puede ser afirmado justificadamente por nosotros» (haciendo así posible la verdad sobre los objetos inexistentes), pero luego pasan a describir la «imagen» como una relación no intencional que presenta un principio de Arquímedes por referencia al cual podemos decir que nuestra teoría actual del mundo, aunque, por supuesto, *verdadera,* puede que no refleje el mundo tan adecuadamente como alguna teoría posterior. La diferencia de terminología carece de importancia, pues lo que quieren los tres filósofos es simplemente hacer que sea posible responder a la pregunta «¿Qué es lo que justifica que nuestras cambiantes teorías del mundo vayan cambiando a mejor en vez de a peor?» Los tres desean una teoría wittgensteiniana del significado-como-uso para resolver lo que yo he denominado problemas de la filosofía «pura» del lenguaje, y una relación figurativa tractaria para resolver los problemas epistemológicos.

La crítica de Putnam a sus propios intentos previos de dar sentido a esta garantía transcendental sirve igual de bien para Sellars y Rosenberg. Dice:

> El realismo metafísico se viene abajo precisamente en el punto en que afirma que es distinguible del realismo peirceano —es decir, de la afirmación de que hay una teoría ideal... Como el propio Peirce (y los verificacionistas) han *dicho* siempre que el realismo metafísico incurre en incoherencia en ese *mismo* punto, y los realistas como yo pensábamos que estaban *equivocados,* no es posible evitar la desagradable admisión de que «ellos tenían razón y nosotros estábamos equivocados» al menos en una cuestión sustancial [29].

Compárese este pasaje con lo que dice Sellars hablando de Peirce:

> ... aunque los conceptos de «verdad ideal» y de «lo que existe realmente» se definen en términos de una estructura conceptual peirciana no exigen que haya nunca una comunidad a lo Peirce.

[29] Putnam, «Realism and Reason», pág. 489 (pág. 130 en *Meaning and the Moral Sciences).*

> El propio Peirce tuvo problemas porque, al no tener en cuenta la dimensión del «reproducir en una imagen», no contaba con ningún principio de Arquímedes fuera de la serie de creencias actuales y posibles en cuyo término definir el ideal o límite al que podrían aproximarse los miembros de esta serie [30].

Lo que dice Sellars aquí es que la identificación por parte de Peirce de «la verdad» con «la opinión llamada a ser aceptada en definitiva por todos» hace parecer que la misma existencia de la verdad y la realidad depende de cuestiones tan arriesgadas como la continuación de la raza de las ideas de la Ilustración sobre la investigación racional. Por eso, Sellars quiere reemplazar una forma de ver la investigación humana que considera «llamada a ser aceptada» como una descripción de un proceso causal que lleva a la creación de autorepresentaciones por el universo. Así vemos cómo Rosenberg se hace eco de la última metafísica idealista de Peirce sobre el amor evolutivo:

> Sólo podemos entender nuestras actividades representacionales... redescribiéndolas en términos de los conceptos de una teoría *total* del universo en cuanto sistema físico que, por necesidad natural, desprende subsistemas que a su vez proyectan necesariamente representaciones cada vez más adecuadas del todo. Dicho tajantemente, debemos llegar a ver el universo físico como un sistema físico integrado que «hace crecer conocedores» necesariamente y que, por tanto, llega a reflejarse dentro de sí mismo [31].

Sellars y Rosenberg tienen razón cuando opinan que la aparición del Espejo de la Naturaleza fue posibilitada por la existencia de mentes, pero insisten (como Putnam) en que la mentalidad y la intencionalidad son irrelevantes para la comprensión de cómo refleja el Espejo. La forma fundamental de representación —aquella que nos ayuda a decir cómo y por qué somos superiores a nuestros antepasados— es la que tiene lugar no en relación con un esquema de convenciones, ni en la relación con las intenciones: «La imagen es una relación compleja y de carácter factual y, en cuanto tal, está en un reducto completamente diferente al de los conceptos de denotación y de verdad» [32].

---

[30] Wilfrid Sellars, *Science and Metaphysics* (Londres y Nueva York, 1968), página 142.

[31] Jay Rosenberg, *Lingustic Representation* (Dordrecht, 1974), pág. 144.

[32] Sellars, *Science and Metaphysics,* pág. 136.

Por eso, si la retractación que hace Putnam es correcta, repercute directamente en el punto que Sellars y Rosenberg consideraban crucial. Putnam está diciendo que el intento de conseguir un conjunto de relaciones no intencionales (como las ofrecidas por una teoría causal de la referencia o por una noción sellarsiana de la «imagen más adecuada») está siempre viciado por el hecho de que dichas relaciones no son más que otras partes de la teoría del mundo de nuestros días. Esta crítica a cualquier posible naturalización de la epistemología sólo nos deja, piensa Putnam, lo que él llama «realismo interno» —la opinión de que podemos explicar «el hecho material de que el uso del lenguaje contribuye a conseguir nuestros objetivos, a recibir satisfacción o a cosas por el estilo» diciendo «no que el lenguaje refleja al mundo, sino que los *hablantes* reflejan el mundo —es decir, su entorno— en el sentido de *construir una representación simbólica de ese entorno*» [38]. El realismo interno, en este sentido, no es más que la concepción de que, según nuestras propias convenciones representacionales, estamos representando el universo mejor que nunca. Pero, a su vez, eso es felicitarnos, por ejemplo, por haber inventado el término *litio* para representar al litio, que no se había representado en todos esos años. La diferencia entre el realismo «metafísico» retractado y el realismo interno no polémico es la diferencia entre decir que estamos representando con éxito atendiendo a las convenciones de representación de la propia Naturaleza y decir que estamos representando con éxito atendiendo a las nuestras. Es la diferencia entre la ciencia como Espejo de la Naturaleza y como conjunto de diagramas de trabajo para hacer frente a la Naturaleza. Decir que nos estamos valiendo bastante bien es cierto, pero superficial. Decir que estamos reflejando correctamente es «sólo una imagen», y además una imagen a la que nunca hemos conseguido dar sentido. Por lo que nosotros podemos saber, la Naturaleza puede hacer aparecer conocedores que la representen, pero no sabemos qué significaría que la Naturaleza sintiera que nuestras convenciones de las representaciones se están pareciendo cada vez más a las suyas y, por tanto, que hoy día está siendo representada más exactamente que en el pasado. O, más bien, sólo podemos ver sentido en esto si seguimos a los Idealistas Absolutos hasta el final, y admitimos que el realismo epistemológico ha de estar basado en el panteísmo personalista.

En esta sección he tratado de presentar la afirmación, simple pero devastadora, de Putnam de que las relaciones no intencionales son

[38] Putnam, «Realism and Reason», pág. 483 (pág. 123 en *Meaning and the Moral Sciences).*

tan relativas a la teoría como las relaciones intencionales en cuanto crítica general a todo el intento de naturalizar la epistemología, transformándola, en primer lugar, en filosofía del lenguaje y luego consiguiendo una explicación naturalista del significado y de la referencia. El motivo común a la «Epistemología Naturalizada» de Quine, a las insinuaciones de Daniel Dennett sobre una «epistemología evolutiva», a la revitalización por Kripke y Fisk de las ideas aristotélicas de esencia y necesidad natural, a varias teorías causales de la referencia y a las teorías sellarsianas de la imagen ha sido des-transcendentalizar la epistemología y al mismo tiempo hacerle realizar lo que siempre habíamos esperado que pudiera hacer: decirnos por qué nuestros criterios de una investigación provechosa no son sólo *nuestros* criterios, sino también los criterios *acertados,* los criterios de la naturaleza, los criterios que nos llevarán a *la verdad*. Si se abandona definitivamente este motivo, la filosofía del lenguaje es simplemente semántica davidsoniana «pura», una semántica que no depende de las imágenes del espejo, sino que, por el contrario, hace tan difícil como posible plantear preguntas filosóficamente interesantes sobre el significado y la referencia.

Por consiguiente, voy a pasar a describir cómo une Davidson su examen de la verdad con su ataque a la distinción «esquema-contenido» y a las imágenes del espejo en general. En primer lugar, trata de decir que la idea de que las afirmaciones sean verdaderas porque corresponden a («reflejan», «representan adecuadamente») la realidad es correcta para todos aquellos casos en que no sea filosóficamente discutible —casos como «La nieve es blanca». Es válida incluso para casos como «La perseverancia mantiene el brillo del honor», «Nuestra teoría del mundo corresponde a la realidad física» y «Nuestra filosofía moral se conforma a la Idea del Bien». Esto es verdad si, y sólo si, el mundo contiene las debidas clases de cosas, y está organizado en la forma que sugieren las afirmaciones. En opinión de Davidson no hay ninguna oportunidad para el «puritanismo filosófico» que dejaría al mundo sin honor, sin realidad física o sin la Idea del Bien. Si se quiere decir que no existen estas cosas, se puede presentar una teoría alternativa del mundo que no las contenga, pero ya no se tratará de una teoría *semántica.* Las discusiones de cómo la verdad es correspondencia con la realidad no dependen de las discusiones de lo que hay en los cielos y en la tierra. No hay caminos que lleven del proyecto de ofrecer las condiciones de verdad de las oraciones del idioma inglés (el inglés tal como se habla, que contiene toda clase de teorías sobre cosas de todas clases) a los criterios para la elección de teorías o a la

construcción de una notación canónica que «dibuje la estructura verdadera y última de la realidad». Para Davidson, la correspondencia es una relación que no tiene preferencias ontológicas —puede unir cualquier tipo de palabra con una cosa de cualquier clase. Esta neutralidad es manifestación del hecho de que, según la concepción davidsoniana, la naturaleza no tiene una forma predilecta de ser representada y, por tanto, no tiene ningún interés en una notación canónica. Tampoco es posible que la naturaleza sea correspondida mejor o peor, a no ser en el sentido elemental de que podemos tener más o menos creencias verdaderas [34].

---

[34] No estoy seguro de la postura de Davidson en el tema de la verdad de las ficciones, ni de si admitiría que se da una relación de «satisfacción» entre «Sherlock Holmes» y Sherlock Holmes. Supongo que sí, pues con ello se subrayaría la separación entre semántica tarskiana y epistemología «realista» en que estoy insistiendo. Según el punto de vista que estoy adoptando, «Sherlock Holmes vivía con el Dr. Watson» es tan verdadero y tiene tan poca necesidad de «análisis filosófico» como «La nieve es blanca». Esto quiere decir que puede haber afirmaciones verdaderas que no contengan expresiones referentes. Esta consecuencia no resulta inquietante si se recuerda la distinción entre «referirse» y «hablar de», y, por tanto, la distinción entre «no referirse» y «no hablar de nada». Pero en las cuestiones relacionadas con la interpretación de Tarski de la idea de «satisfacción» no tengo las cosas tan claras como para estar seguro de la relación entre esta idea y una de las otras dos. La línea general que voy a seguir es que hay afirmaciones verdaderas sobre las ficciones, valores y números, además de sobre los gatos que están en una alfombra, y que el intento de encontrar algo parecido a «correspondencia» en términos de lo cual «analizar» las primeras verdades siguiendo el modelo de la última no tiene sentido. Sellars señala este punto diciendo que no todas las afirmaciones verdaderas «pintan» el mundo —sólo las «empíricas básicas». Yo preferiría decir que ninguna afirmación verdadera es una pintura del mundo —que pintar es «sólo una imagen» y, una imagen que sólo ha servido para producir *Sprachstreit* todavía más retorcidos.

Sobre la interpretación de Tarski, véase la afirmación de Sellars de que «las afirmaciones semánticas de la línea Tarski-Carnap no afirman relaciones entre elementos lingüísticos y extra-lingüísticos» (*Science and Metaphysics,* página 82), y el contraste entre «el sentido 'gráfico' de 'corresponder' y el sentido de 'corresponder' en Tarski-Carnap» (pág. 143). La opinión de Sellars es que las afirmaciones semánticas versan sobre intensiones, y que el «pintar» no tiene nada que ver con la semántica. Véanse posiciones encontradas en John Wallace, «On the Frame of Reference», en *Semantics of Natural Language,* edición Davidson y Jarman, y Hartry Field, «Tarski's Theory of Truth», *Journal of Philosophy,* 69 (1972).

Véase Robert Brandom, «Truth and Assertibility», *Journal of Philosophy,* 73 (1976), pág. 137-149, donde intenta desarrollar una concepción de «verdadero» que lo distinga de «afirmable» *sin* tratar de interpretarlo como «que está en relación de correspondencia con lo extra-lingüístico». Brandom presenta una explicación de por qué necesitamos «verdadero» además de «afirmable» con vistas a lo que yo estoy llamando filosofía «pura» del lenguaje —entender

En segundo lugar, Davidson piensa que la idea de «esquema representacional» o «marco de referencia conceptual» o «correspondencia intencionada» trata de separar la idea de «verdad» de la de «significado y, por tanto, tiene que fracasar. Su forma más eficaz de expresar esto aparece en su afirmación de que la idea de un «esquema conceptual alternativo» —por ejemplo, un esquema que no contuviera ninguna de las expresiones referentes usadas en el nuestro— es la idea de un lenguaje que sea «verdadero pero no traducible». Tras una crítica constante a algunas variaciones sobre las tradicionales imágenes del espejo (esquemas conceptuales como el «encajar) la realidad meojr o peor, o «clasificar» la realidad de distinta manera unos de otros), Davidson concluye:

> El problema es que la idea de encajar con la totalidad de la experiencia, como las ideas de encajar con los hechos o de ser conforme a los hechos, no añade nada inteligible al concepto elemental de ser verdadero. Hablar de experiencia sensorial en vez de hablar de testimonios, o simplemente de hechos, expresa una concepción sobre la fuente o naturaleza de los testimonios, pero no añade una nueva realidad con la cual se puedan contrastar los esquemas conceptuales.

Por eso, dice:

> Nuestro intento de describir los lenguajes o los esquemas conceptuales en términos de la noción de encajar con alguna realidad ha venido a reducirse al simple pensamiento de que algo es una teoría o esquema conceptual aceptable si es verdadero. Quizá deberíamos decir *en gran parte* verdadero para dar lugar a que los que comparten un sistema difieran en los detalles. Y el criterio de un esquema conceptual diferente del nuestro se convierte en gran parte verdadero pero no traducible. La pregunta de si este criterio es útil es lo mismo que si preguntamos por la mayor o menor perfección con que entendemos la idea de verdad, cuando se aplica al lenguaje, independientemente de la idea de traducción. La respuesta es, creo yo, que no la entendemos independientemente. En absoluto [35].

---

cómo funciona un lenguaje, en oposición a cómo está conectado con el mundo. Por eso diagnostica, creo yo, la confusión fundamental entre la necesidad de este concepto en semántica y su necesidad en epistemología, confusión que motiva la filosofía «impura» del lenguaje.

[35] Davidson, «On the Very Idea of a Conceptual Scheme», pág. 16. Por desgracia, Davidson malinterpreta a Kuhn y piensa que quiere decir «intraducible» cuando dice «inconmensurable» (pág. 12). Dentro de la argu-

La razón por la que Davidson dice que no entendemos esto procede de lo que él llama una «cierta concepción holística del significado»:

> Si las oraciones dependen, en cuanto a su significado, de su estructura, y nosotros entendemos el significado de cada ítem de la estructura únicamente como una abstracción de la totalidad de las oraciones en que figura, podemos presentar el significado de cualquier oración (o palabra) únicamente presentando el significado de cada una de las oraciones (y palabras) de la lengua [36].

Esta concepción holística del significado equivale a la idea de que una teoría del significado aplicada a una lengua *lo único* que puede hacer es «ofrecer una explicación de cómo los significados de las oraciones dependen de los significados de las palabras» (pág. 304). El paso fundamental consiste en decir que no es necesario pensar que «las palabras individuales tengan significados en absoluto, en ningún sentido que transcienda el hecho de que tienen una influencia sistemática sobre los significados de las oraciones en que ocurren» (página 305). La concepción tradicional es que anclamos el lenguaje en el mundo dando significado por ostensión (o algún otro mecanismo no intencional) a ciertas palabras, y luego procediendo holísticamente a partir de ese punto. La idea de Davidson y de los neo-wittgensteinianos es que incluso «rojo» y «mamá» tienen usos —pueden ayudar a hacer posible la formulación de verdades— únicamente en el contexto de oraciones y, por tanto, de todo un lenguaje (cfr. página 308). Cualquiera que sea el papel que desempeña la ostensión (o las vías nerviosas, o cualquier otra construcción no intencional) en el aprendizaje de un lenguaje, no hace falta saber nada sobre estos mecanismos para saber la lengua, ni saber cómo se traduce dicha lengua. Lo mismo vale, señala Davidson, para las condiciones de verdad traskianas en el caso de las oraciones en inglés, y ésta es la razón por la que resulta que «una teoría del significado para la lengua L hace ver 'cómo los significados de las oraciones dependen de los significados de las palabras' si contiene una definición (recursiva) de la verdad-en-L», pues

> lo que hace falta para una teoría del significado en una lengua L es que sin recurrir a (otras) nociones semánticas establezca sobre el predicado 'es T' [en el esquema: *s* es T si y sólo si p] las

mentación del presente libro es muy importante distinguir estas dos nociones. Véase el capítulo séptimo, sección 1.

[36] Davidson, «Truth and Meaning», pág. 308.

> restricciones suficientes como para vincular a todas las oraciones procedentes del esquema T cuando se sustituye a 's' por una descripción estructural de una oración de L y a 'p' por esa oración... y, por tanto, la condición que hemos impuesto a las teorías satisfactorias del significado es en esencia la Convención T de Tarski (pág. 309).

Por eso, Davidson nos invita a concluir:

> Dado que la Convención T encarna nuestra mejor intuición sobre la forma en que se usa el concepto de verdad, parece que no se deben tener muchas esperanzas en una prueba de que un esquema conceptual es radicalmente diferente del nuestro si esa prueba depende de la suposición de que podemos separar la idea de verdad de la de traducción [87].

Voy a permitirme reformular el fondo de este argumento señalando la conexión entre el holismo de Davidson sobre el significado y su desdén por la noción de «esquema». El que no sea holista en este sentido pensará que la comprensión de una lengua consiste en dos procesos distintos —unir algunas palabras individuales con el mundo a través de la ostensión, y luego dejar que otras palabras vayan incorporando significados en torno a este núcleo central en el curso de su utilización. Pensará también que entender lo que significa «verdad» implica «analizar» cada oración verdadera hasta que aparezcan ostensiones que la hagan verdadera. Esta imagen del holismo que deja de aplicarse en el punto en que la referencia es menos problemática —en la superficie de contacto entre la lengua y el mundo donde los demostrativos realizan su tarea— es una forma de mantener en pie la distinción esquema-contenido. Si pensamos así del lenguaje, se nos ocurrirá pensar que otros (los galácticos, por ejemplo) habrán «fragmentado» el mundo de otra manera en sus actos originales de ostensión y, por tanto, habrán dado diferentes significados a las palabras individuales que forman el «núcleo» de su lengua. El resto de ésta se verá contagiada por esta divergencia de nuestra forma de dar significado al «núcleo» del inglés y, por tanto, no tendremos ninguna forma de comunicación —ningún punto común de referencia, ninguna posibilidad de traducción. El ataque de Davidson a estas metáforas es paralelo a la crítica que el wittgensteiniano Max Black hace del problema llamado del espectro invertido —la posibilidad de que algunas personas puedan (en el caso de los colores y quizá de todos los demás

[87] Davidson, «On the Very Idea of a Conceptual Scheme», pág. 17.

«espacios cualitativos» perceptivos) entrar con mal pie. Black señala que podemos hacer una división, gracias a esta diferencia, para todas las cuestiones de comunicación —el lenguaje conseguirá los mismos resultados independientemente de lo que estuviéramos mostrando [38]. De la misma manera, Davidson puede decir que si una diferencia en la ostensión original no se manifiesta en el nivel holístico —en el uso de las oraciones que contienen la palabra— la teoría del significado para la lengua puede hacer una división atendiendo a esa diferencia.

Sin embargo, podríamos querer insistir en que el argumento de Davidson sólo se refiere a lo que es presentar una *teoría* del significado y una *teoría* de la verdad. Podríamos decir que lo único que podría hacer es demostrar que no podemos *verificar* la existencia de una lengua que ofrezca una descripción verdadera del mundo a no ser que se pueda traducir a la nuestra, y que esto no demostraría que no puede *haber* una. Esta línea de argumentación sería análoga a la crítica a la forma en que Black se ocupa del problema del «espectro invertido», y a estrategias wittgensteinianas antiescépticas semejantes, tachándolas de «verificacionistas» [39]. Concluiré este capítulo respondiendo a esta objeción.

6. Verdad, bondad y relativismo

Para alcanzar una visión más amplia del problema del verificacionismo, consideremos las siguientes propuestas:

> Ninguno de nuestros términos hace referencia.
> Ninguna de nuestras creencias es verdadera.

---

[38] Véase Max Black, «Linguistic Method in Philosophy», en *Language and Philosophy* (Ithaca, 1949), especialmente págs. 3-8. Black trata el «espectro invertido» como paradigma del tipo de argumento escéptico que el enfoque del lenguaje que él recomienda contribuye a desacreditar. Creo que en esto tien razón. En realidad, hay una perspectiva a partir de la cual todo lo que he llamado «conductismo epistemológico» en el capítulo cuarto puede hacerse proceder del tipo de argumento que Black presenta aquí.

[39] Sobre este tipo de objeción véase Barry Stroud, «Transcendental Arguments», *Journal of Philosophy,* 65 (1968), págs. 241-256. He tratado de hacer frente a esta objeción en «Verificationism and Transcendental Arguments», *Nous,* 5 (1971), 3-14, y a algunas objeciones relacionadas con la anterior en «transcendental Arguments, Self-Reference and Pragmatism», en *Transcendental Arguments and Science,* ed. Peter Bieri, Rolf Horstmann y Lorenz Krüger (Dordrecht, 1979), pág. 77-103.

No es posible ninguna traducción entre el inglés y el lenguaje que representa cómo es el mundo.
Ninguna de nuestras intuiciones morales es correcta (participa en la Idea del Bien, refleja la Ley Moral, o algo por el estilo).

Creo que la menos plausible de todas es la última. La razón es que lo que consideramos que es moral —que es incluso *aspirante* a la participación en la Idea del Bien— es algo que satisface, o al menos no resulta demasiado incompatible con, nuestras intuiciones morales actuales. ¿Nos comportamos irreflexivamente al pensar que «moralmente correcto» significa «algo parecido a nuestro propio ideal de conducta, pero más aún»? ¿Significa esto que despreciamos una posibilidad real simplemente por la intuición verificacionista de que no podemos reconocer a nadie que *no* atraiga a nuestras intuiciones porque tenga algo que decir sobre la moralidad?

Creo que la mejor forma de contestar a estas preguntas es distinguiendo entre un sentido filosófico de «bueno», en relación con el cual se presenta «la falacia naturalista», y un sentido corriente, en relación con el cual no ocurre eso. Si preguntamos: «¿Cuáles son las condiciones necesarias y suficientes para que algo sea bueno?», podemos decir, con G. E. Moore, que siempre es una cuestión abierta saber si «bueno» tiene algo que ver con cualquier condición que se pueda mencionar. Las palabras como «bueno», *una vez que se han manejado en la forma en que las ha manejado la tradición filosófica,* adquieren un sentido en relación con el cual la respuesta sería afirmativa. Se convierten en el nombre de un *focus imaginarius,* una Idea de la Razón Pura cuyo sentido total no es ser identificable con cualquier conjunto de condiciones. Esto no equivale a decir que ἀγαθός sirviera a los griegos pre-platónicos como nombre de algo tan incondicionado, ni que hubiera ninguna palabra que sirviera para ello antes de que los pitagóricos y los órficos y Platón inventaran el idealismo (en sus dos sentidos). Pero ahora hay un uso específicamente filosófico de «bueno», un uso que no sería el que es si no fuera porque Platón, Plotino, Augstín y otros han ayudado a construir una teoría específicamente platónica de la diferencia absoluta entre lo eterno y lo espacio-temporal. Cuando Moore apeló a nuestro sentido sobre el significado de «bueno» sólo estaba recurriendo a este sentido, sentido que es bastante difícil de captar sin tener conocimiento de la historia del pensamiento occidental, más común en Bloomsbury que en Birmingham. *Principia Ethica* debería leerse, lo mismo que *Euthyphro,* como una obra de pedagogía edifi-

cante y de tranquila revolución moral, más que como un esfuerzo por describir la práctica intelectual o lingüística en boga.

Pero existe también un sentido corriente de «bueno», el sentido que tiene la palabra que se usa para alabar —hacer notar que algo responde a algún interés. En este sentido, no nos vamos a encontrar con un conjunto de condiciones necesarias y suficientes para la bondad que nos permitan encontrar la Buena Vida, resolver los dilemas morales, clasificar las manzanas, etc. Son demasiados los tipos de intereses diferentes que atender, demasiados los tipos de cosas que recomendar y demasiadas las razones diferentes para recomendarlas, para que se pueda encontrar semejante conjunto de condiciones necesarias y suficientes. Pero ésta es una razón muy diferente de la in-definibilidad del sentido filosófico de «bueno». En su sentido llano y trasnochado, la razón por la que «bueno» es indifinible no es que podamos estar totalmente equivocados sobre lo que son hombres buenos o manzanas buenas, sino simplemente que *ningún* término descriptivo interesante tiene condiciones interesantes necesarias y suficientes. En el primer sentido de «bueno», el filosófico, el término es indefinible porque cualquier cosa que digamos sobre lo que es bueno puede ser «lógicamente» irrelevante para lo que es la bondad. La única forma de que una mente sencilla y trasnochada capte este primer sentido es comenzar con Platón o Moore y esperar que se haga con la Idea.

Lo que me interesa al decir que hay dos sentidos de «bueno» es, ciertamente, hacer plausible la sugerencia de que hay también dos sentidos de «verdadero» y «real» y «representación correcta de la realidad», y que la mayoría de las perplejidades de la epistemología proceden de la vacilación entre ellos (lo mismo que la mayoría de las perplejidades de la meta-ética proceden de la vacilación entre los sentidos de «bueno»). Si comenzamos analizando la analogía entre bondad y verdad, debemos tener en cuenta el uso corriente de «verdadero», con el significado aproximado de «lo que puedes defender frente a cualquiera que se presente». En este caso, la línea de separación entre que una creencia esté justificada y sea verdadera es muy fina. Esa es la razón por la que Sócrates tenía problemas para explicar la diferencia entre estas dos nociones, el mismo problema con que los profesores de fisolofía nos encontramos al explicársela a los estudiantes de primer curso. Cuando los pragmatistas identificaban la verdad con «lo que creeremos si seguimos investigando con nuestras luces actuales» o «lo que es mejor que creamos» o con «afirmabilidad justificada», se creían que estaban siguiendo los pasos de Mill, y haciendo por la ciencia lo que los

utilitaristas habían hecho por la moralidad —convertirla en algo que se podía utilizar en vez de algo que sólo se podía respetar, algo que tenía una continuidad con el sentido común en vez de algo que podía quedar tan alejado del sentido común como la Mente de Dios.

Tarski y Davidson se fijan en el sentido llano y desfasado de «verdadero», mientras que Putnam aplica su argumento de la «falacia naturalista» al sentido filosófico especial[40]. Las dos actitudes se ignoran mutuamente. Davidson representa el proyecto «puro» de encontrar un sistema para representar en forma de diagrama las relaciones entre las oraciones del inglés para así dejar claro por qué las personas dicen que algunas oraciones largas son «verdaderas» por el hecho de ver esto en función de que llaman «verdaderas» a otras oraciones más cortas. Putnam representa el proyecto «impuro» de hacernos ver que la comprensión más completa posible de esta materia nos dejará igual de abiertos que siempre a la posibilidad de que no tengamos creencias verdaderas en absoluto. Es cierto, pero, una vez más, se da una ambigüedad que hace que la irrelevancia parezca oposición. Davidson puede decir, sin intentar en absoluto prever el fin de la investigación y bloquear la «meta-inducción» escéptica de Putnam, que la mayoría de nuestras creencias son verdaderas. Esta afirmación se deduce de la afirmación también suya de que no podemos entender la sugerencia de que la mayoría de ellas son falsas —sugerencia que sólo tiene sentido cuando está respaldada por la idea absurda de un «esquema conceptual alternativo e intraducible». Pero no tiene ninguna fuerza contra el escéptico, a no ser en la medida en que le desafía a concretar su indicación de que lo hemos entendido todo al revés —proyecto tan difícil como sería el de concretar la afirmación más limitada de que todas nuestras oraciones que terminan en «... es moralmente correcto» son falsas. Davidson, como Kant, está diciendo que del hecho de que esta o aquella creencia central resulte falsa, o de que alguna intuición moral sea perversa y fruto de los prejuicios, no se puede deducir la posibilidad de que ocurra lo mismo con *todas* ellas. La duda sobre su verdad o bondad sólo tiene sentido en el contexto de un acuerdo general. El escéptico y Putnam escurren el bulto cambiando al sentido específicamente «filosófico» de «bueno» y «verdadero» que, como las Ideas de la Razón Pura, están concebidos precisamente para aspirar a la categoría de Incondicionado —lo que

[40] Véase «Reference and Understanding» *(Meaning and the Moral Sciences,* página 108), donde Putnam explicita la analogía entre su estrategia con «verdadero» y la de Moore con «bueno».

escapa del contexto en que se realiza el discurso y se lleva a cabo la investigación, y trata de establecer un nuevo contexto.

Davidson parece verificacionista y convencionalista y relativista si su afirmación de que «la mayoría de las creencias son verdaderas» se interpreta como que estamos llegando al final de la indagación, a punto de captar cómo es el mundo, de conseguir un Espejo de la Naturaleza perfectamente limpio. Pero esto es igual que interpretarle como si dijera que estamos cerca de la cumbre de la línea dividida de Platón, de una visión clara de la Idea del Bien. Tal interpretación considera que «verdadero» y «bueno» son nombres que se aplican a la adaptación exacta de un esquema de representaciones a algo que proporciona contenido a tal esquema. Pero como dice Davidson:

> Sería un error resumir esto diciendo que hemos demostrado cómo es posible la comunicación entre personas que tienen esquemas diferentes, algo que funciona sin necesidad de lo que no puede haber, a saber, un terreno neutral, o un sistema coordinado común. No hemos encontrado ninguna base inteligible que nos permita decir que los esquemas son diferentes. Sería igualmente erróneo proclamar la gran noticia de que toda la humanidad —todos los hablantes de un idioma, al menos— tienen una ontología y un esquema comunes. Si no podemos decir de forma inteligible que los esquemas son diferentes, tampoco podemos decir de forma inteligible que sean uno.
>
> Al abandonar la dependencia del concepto de una realidad no interpretada, de algo exterior a todos los esquemas y a la ciencia, no renunciamos a la idea de verdad objetiva —sino todo lo contrario—. Dado el dogma de un dualismo de esquema y realidad, conseguimos una relatividad conceptual y una verdad relativa a un esquema. Sin el dogma, esta forma de relatividad se va al traste. Como es natural, la verdad de las oraciones sigue siendo relativa al lenguaje, pero esto es algo perfectamente objetivo. Al abandonar el dualismo de esquema y mundo, no renunciamos al mundo, sino que restablecemos el contacto no mediado con los objetos conocidos cuyas travesuras hacen que nuestras oraciones y opiniones sean verdaderas o falsas [41].

Decir que Davidson es verificacionista y relativista cuando dice que la mayoría de nuestras creencias son verdaderas o que cualquier

[41] Davidson, «On the Very Idea of a Conceptual Scheme», pág. 20. He intentado desarrollar el contraste entre abandonar el mundo (tal como lo considera el sentido común) y abandonar el Mundo (en cuanto la Cosa-en-Sí, el Mundo que quizá nunca lleguemos a captar correctamente) en «The World Well Lost», *Journal of Philosophy*, 69 (1972), 649-666, especialmente 662-663.

idioma se puede traducir al inglés, es lo mismo que si dijéramos que no está utilizando las nociones «platónicas» de Verdad y Bondad y Realidad que necesitan los «realistas» para hacer que su realismo sea más llamativo y controverdido (más «metafísico» que «interno», en el sentido de Putnam). Pero tampoco está «refutando» Davidson estas ideas platónicas al demostrar su «incoherencia». Lo único que puede hacer con ellas es lo que hizo Kant con las Ideas de la Razón Pura —hacer ver cómo funcionan, qué pueden y no pueden hacer. El problema de las ideas platónicas no es que estén «equivocadas», sino que no hay mucho que decir sobre ellas —en concreto, no hay forma de «naturalizarlas» o conectarlas de alguna otra manera con el resto de la investigación, o la cultura o la vida. Si preguntamos a Dewey por qué piensa que la cultura occidental tiene la más ligera idea de qué es la bondad, o a Davidson por qué cree que hablamos alguna vez sobre lo que existe realmente o decimos algo verdadero al respecto, es probable que nos pregunten qué es lo que nos hace tener dudas sobre ello. Si respondemos que son ellos quienes deben encargarse de presentar pruebas, y que no les está permitido argumentar partiendo del hecho de que no lo sabríamos nunca si el escéptico estuviera en lo cierto para llegar a la conclusión de que es imposible que tenga razón, Dewey y Davidson podrían responder que *no* argumentan de esa manera. No tienen necesidad de invocar argumentos verificacionistas; sólo tienen que preguntar por qué deben preocuparse por la alternativa escéptica mientras no se les presente alguna razón concreta para dudar. Decir que este intento de hacer recaer la carga de la prueba sobre el «verificacionismo» escéptico, o una confusión del orden del conocimiento con el orden del ser, es como llamar «verificacionista» a quien diga que no se molestará en saber si las cosas que él ha llamado «rojas» son realmente rojas mientras no se presente una alternativa concreta. La decisión sobre si deben tenerse criterios más altos de lo normal para la aplicación de palabras como «verdadero» o «bueno» o «rojo» no es, en mi opinión, una cuestión discutible. Pero sospecho que es el único problema que queda entre realistas y pragmatistas, y estoy casi seguro de que la filosofía del lenguaje no nos ha brindado nuevos puntos controvertidos interesantes.

TERCERA PARTE

La Filosofía

Capítulo VII

De la Epistemología a la Hermenéutica

1. Conmensuración y conversación

Ya he indicado (en el capítulo tercero) que el deseo de una teoría del conocimiento es un deseo de constricción —un deseo de encontrar «fundamentos» a los que poder agarrarse, armazones que no nos dejen extraviarnos, objetos que se impongan a sí mismos, representaciones que no se puedan negar. Cuando describí la reacción actual contra la búsqueda de fundamentos dándole el nombre de «conductismo epistemológico» (en el capítulo cuarto), no quería dar a entender que Quine y Sellars nos permitirían tener una clase de epistemología que fuera «conductista», nueva y mejor. Más bien, nos hacen ver qué aspecto tienen las cosas cuando renunciamos al deseo de confrontación y constricción. Sin embargo, muchas veces se tiene la sensación de que la desaparición de la epistemología en cuanto búsqueda de fundamentos deja un vacío que hay que cubrir. En los capítulos quinto y sexto he criticado varios intentos realizados para llenar ese hueco. En el presente capítulo me ocuparé de la hermenéutica, por lo que desde el primer momento quiero dejar en claro que *no* estoy presentando a la hermenéutica como «sucesora» de la epistemología, como una actividad que ocupe el vacío cultural ocupado en otros tiempos por la filosofía centrada epistemológicamente. En la interpretación que voy a presentar, «hermenéutica» no es el nombre de una disciplina, ni de un método de conseguir los resultados que la epistemología no consiguió obtener, ni de un programa de investigación. Por el contrario, la hermenéutica es una expresión de esperanza de que el espacio cultural dejado por el abandono de la epistemología no llegue a llenarse —que nuestra cultura sea una cultura en la que ya no se siente la exigencia de

constricción y confrontación. La idea de que un armazón neutro y permanente cuya «estructura» puede mostrar la filosofía es la idea de que los objetjos que van a ser confrontados por la mente, o las reglas que constriñen la investigación, son comunes a todo discurso, o al menos a todo discurso que verse sobre un tema determinado. Así, la epistemología avanza partiendo de la suposición de que todas las aportaciones a un discurso determinado son conmensurables. La hermenéutica es en gran parte una lucha contra esta suposición.

Por «conmensurable» entiendo que es capaz de ser sometido a un conjunto de reglas que nos dicen cómo podría llegarse a un acuerdo sobre lo que resolvería el problema en cada uno de los puntos donde parece haber conflicto entre las afirmaciones [1]. Estas reglas nos indican cómo se construye una situación ideal, en la que todos los desacuerdos residuales se considerarán «no cognitivos» o meramente variables, o de lo contrario meramente temporales —capaces de ser resueltos haciendo algo más. Lo que importa es que debe haber acuerdo sobre lo que habría que hacerse si se tuviera que llegar a tomar una resolución. Mientras tanto, los interlocutores pueden estar en acuerdo en diferir —estando satisfecho cada uno de la racionalidad del otro. La idea dominante de la epistemología es que para ser racional, para ser plenamente humano, para hacer lo que debemos, hemos de ser capaces de llegar a un acuerdo con otros seres humanos. Construir una epistemología es encontrar la máxima cantidad de terreno que se tiene en común con otros. La suposición de que se puede construir una epistemología es la suposición de que ese terreno existe. Algunas veces se ha imaginado que este terreno común está fuera de nosotros —por ejemplo, en el mundo del Ser en oposición al del Devenir, en las Formas que orientan la investigación y al mismo tiempo son su meta. Otras Otras veces se ha imaginado que están dentro de noostros, como en la idea del siglo XVII de que al entender nuestras propias mentes debemos ser capaces de entender el método adecuado para encontrar la verdad. Dentro de la filosofía analítica, muchas veces se ha imaginado que estaba en el lenguaje, del que se suponía que proporcionaba el esquema universal para todo posible contenido. Insinuar

---

[1] Téngase en cuenta que este sentido de «conmensurable» no es el mismo de «atribuir el mismo significado a los términos». Este sentido —que es el que se utiliza muchas veces al hablar de Kuhn— no me parece útil, dada la fragilidad de la idea de «igualdad de significado». Decir que las partes que intervienen en una controversia «utilizan los términos de distintas maneras» me parece una forma poco reveladora de describir el hecho de que no pueden encontrar una forma de ponerse de acuerdo en qué es lo que cerraría la discusión. Véase capítulo 6, sección 3, sobre este punto.

que *no* existe este terreno común parece que es poner en peligro la racionalidad. Cuestionar la necesidad de conmensuración parece el primer paso hacia la reanudación de la guerra de todos contra todos. Así, por ejemplo, una reacción común frente a Kuhn y Feyerabend es decir que están proponiendo el uso de la fuerza en vez de la persuasión.

Los tratamientos del conocimiento holistas, antifundacionalistas y pragmatistas que encontramos en Dewey, Wittgenstein, Quine, Sellars y Davidson son casi igual de ofensivos para muchos filósofos, precisamente porque abandonan la búsqueda de conmensuración y, por tanto, son «relativistas». Si negamos que haya fundamentos que sirvan como base común para juzgar las pretensiones de conocimiento, parece ponerse en peligro de idea del filósofo como guardián de la racionalidad. En un sentido más general, si decimos que no existe la epistemología y que no es posible encontrarle un sustituto, por ejemplo, en la psicología empírica o en la filosofía del lenguaje, puede dar la impresión de que decimos que no existe lo que se llama acuerdo y desacuerdo racional. Las teorías holistas parecen dar autorización a todo el mundo para que construya su propio mundo —su propio paradigma, su propia práctica, su propio juego lingüístico— y luego se deslice en su interior.

Creo que la idea de que la epistemología, o alguna disciplina que la suceda adecuadamente, es necesaria para la cultura confunde dos papeles que puede desempeñar el filósofo. El primero es el del intermediario socrático entre varios discursos. En su tertulia, por así decirlo, se consigue que los pensadores herméticos abandonen sus prácticas encerradas en sí mismas. En el curso de la conversación se llega a compromisos o se transcienden los desacuerdos entre disciplinas y discursos. El segundo papel es el de supervisor cultural, que conoce el terreno común de todos y cada uno —el rey-filósofo platónico que sabe lo que están haciendo realmente todos los demás tanto si lo saben *ellos* como si no, pues tiene conocimiento del contexto último (las Formas, la Mente, el Lenguaje) dentro del que se están haciendo. El primer papel es el adecuado para la hermenéutica, el segundo para la epistemología. La hermenéutica ve las relaciones entre varios discursos como los cabos dentro de una posible conversación, conversación que no presupone ninguna matriz disciplinaria que una a los hablantes, pero donde nunca se pierde la esperanza de llegar a un acuerdo mientras dure la conversación. No es la esperanza en el descubrimiento de un terreno común existente con anteroridad, sino *simplemente* la esperanza de llegar a un acuerdo, o, cuando menos, a un desacuerdo interesante y fructífero.

La epistemología considera la esperanza de llegar a un acuerdo como una señal de la existencia de un terreno común que, quizá sin que lo sepan los hablantes, les une en una racionalidad común. Para la hermenéutica, ser racional es estar dispuesto a abstenerse de la epistemología —de pensar que haya un conjunto especial de términos en que deben ponerse todas las aportaciones a la conversación— y estar dispuestos a adquirir la jerga del interlocutor en vez de traducirla a la suya propia. Para la epistemología, ser racional es encontrar el conjunto adecuado de términos a que deberían traducirse todas las aportaciones para que sea posible el acuerdo. Para la epistemología, la conversación es investigación implícita. Para la hermenéutica, la investigación es conversación rutinaria. La epistemología ve a los participantes unidos en lo que Oakeshott llama una *universitas* —grupo unido por intereses mutuos en la consecución de un fin común. La hermenéutica los ve unidos en lo que él llama una *societas* —personas cuyos caminos por la vida se han juntado, unidas por la urbanidad más que por un objetivo común, y mucho menos por un terreno común[2].

Quizá parezca forzada mi utilización de los términos *epistemología* y *hermenéutica* para designar estos opuestos ideales. Trataré de justificarlo haciendo notar algunas de las conexiones existentes entre holismo y el «círculo hermenéutico». La idea del conocimiento como representación exacta se presta naturalmente a que se entienda como que ciertas clases de representaciones, ciertas expresiones, ciertos procesos son «básicos», «privilegiados» y «tienen carácter de fundamento». La críticas a esta concepción que he examinado en los capítulos anteriores se ven confirmadas por argumentos holísticos de esta forma: no podremos aislar elementos básicos a no ser basándonos en un conocimiento previo de toda la estructura dentro de la que ocurren estos elementos. Por eso, no podremos conseguir que la noción de «representación exacta» (elemento-por elemento) sustituya a la de realización con éxito de una práctica. Nuestra elección de los elementos estará dictada por nuestra comprensión de la práctica, en vez de que la práctica esté «legitimada» por una «reconstrucción racional» a partir de elementos. Esta línea argumental holística dice que nunca seremos capaces de evitar el «círculo hermenéutico» —el hecho de que no podemos entender las partes de una cultura, práctica, teoría, lenguaje o cualquier otra cosa desconocida, a no ser que sepamos algo sobre cómo funciona todo el conjunto, mientras que no podemos captar cómo funciona todo

[2] Cfr. «On the Character of a Modern European State», en Michael Oakeshott, *On Human Conduct* (Oxford, 1975).

el conjunto hasta que no tengamos alguna comprensión de sus partes. Esta idea de la interpretación indica que llegar a entender se parece más a tener conocimiento de una persona que a seguir una demostración. En ambos casos avanzamos y retrocedemos entre distintas opiniones sobre cómo caracterizar las afirmaciones particulares u otros hechos, y opiniones sobre el sentido de toda la situación, hasta que poco a poco nos vamos sintiendo a gusto con lo que hasta ahora nos era extraño. La idea de la cultura como una conversación más que como una estructura levantada sobre unos fundamentos encaja bien con esta idea hermenéutica del conocimiento, pues entrar en conversación con desconocidos es, igual que la adquisición de una nueva virtud o destreza imitando modelos, cuestión de φρόνησις más que de επιστήμη.

La forma habitual de tratar la relación entre hermenéutica y epistemología es sugerir que se dividan la cultura entre sí —la epistemología se ocuparía de la parte «cognitiva» seria e importante (aquélla en que cumplimos nuestras obligaciones con la racionalidad) y la hermenéutica de todo lo demás. La idea que está detrás de esta división es que el conocimiento en sentido estricto —επιστήμη— debe tener un λόγος y que un λόγος sólo puede ser dado por el descubrimiento de un método de conmensuración. La idea de conmensurabilidad está incluida en la noción de «cognición auténtica», por lo que es «sólo cuestión de gusto» o «de opinión» no tiene que caer necesariamente dentro de la responsabilidad de la epistemología, y por el contrario, lo que la epistemología no puede hacer conmensurable se desprecia como meramente «subjetivo».

El enfoque pragmático del conocimiento sugerido por el conductismo epistemológico considerará que la línea divisoria entre los discursos que pueden hacerse conmensurables y los que no es meramente la que hay entre discurso «normal» y «anormal» —distinción que generaliza la distinción de Kuhn entre ciencia «normal» y «revolucionaria». La ciencia «normal» es la práctica de resolver los problemas teniendo el trasfondo de un consenso sobre lo que se tiene por una buena explicación de los fenómenos y sobre lo que haría falta para que un problema estuviera resuelto. La ciencia «revolucionaria» es la introducción de un nuevo «paradigma» de explicación y, por tanto, de un nuevo conjunto de problemas. La ciencia normal está todo lo cerca que puede estar la vida real de la idea del epistemólogo de lo que significa ser racional. Todos están de acuerdo en cómo evaluar todo lo que dicen todos los demás. En sentido más general, el discurso normal es aquel que se realiza dentro de un conjunto admitido de convenciones sobre lo que debe tenerse

por aportación relevante, lo que debe tenerse por respuesta a una pregunta, y lo que debe tenerse por un buen argumento en favor de esa respuesta o una buena crítica contra la misma. El discurso anormal es lo que ocurre cuando interviene en el discurso alguien que ignora estas convenciones o que las deja de lado. La ἐπιστήμη es producto del discurso normal —el tipo de afirmación en que pueden estar de acuerdo que es verdadera todos los participantes a quienes los otros participantes tienen por «racionales». El producto del discurso anormal puede ser cualquier cosa comprendida entre lo absurdo y la revolución intelectual, y no hay ninguna disciplina que lo describa, lo mismo que no hay ninguna disciplina dedicada al estudio de los imprevisibles o de la «creatividad». Pero la hermenéutica es el estudio de un discurso anormal desde el punto de vista de un discurso normal —el intento de dar cierto sentido a lo que está pasando en un momento en que todavía no estamos seguros sobre ello como para hacer una descripción y, por tanto, para comenzar su explicación epistemológica. El hecho de que la hermenéutica suponga, inevitablemente, alguna norma le da cierto carácter «Whiggish». Pero en la medida en que procede no-reductivamente y con la esperanza de descubrir un nuevo ángulo en las cosas, puede transcender ese carácter.

Desde este punto de vista, la línea divisoria entre los respectivos dominios de la epistemología y la hermenéutica no consiste en la diferencia entre las «ciencias de la naturaleza» y las «ciencias del hombre», ni entre hecho y valor, ni entre tórico y práctico, ni entre «conocimiento objetivo» y algo más viscoso y dudoso. La diferencia es cuestión de familiaridad, simplemente. Seremos epistemológicos donde comprendamos perfectamente bien lo que está ocurriendo, pero queremos codificarlo para ampliarlo, fortalecerlo, enseñarlo o «buscarle una base». Tenemos que ser hermenéuticos cuando no comprendamos lo que está ocuriendo, pero tenemos la honradez de admitirlo, en vez de adoptar una actitud *descaradamente* «Whiggish» al respecto. Esto significa que sólo podemos conseguir la conmensuración epistemológica donde ya tenemos prácticas admitidas de investigación (o, más en general, de discurso) —con la misma facilidad en el arte «académico», filosofía, «escolástica» o política «parlamentaria» que en la ciencia «normal». Podemos conseguirla no porque hayamos descubierto algo sobre «la naturaleza del conocimiento humano», sino simplemente porque cuando una práctica lleva el tiempo suficiente las convenciones que la hacen posible —y que permiten un consenso sobre cómo dividirla en partes— se pueden asilar con relativa facilidad. Nelson Goodman ha dicho de la

inferencia inductiva y deductiva que descubrimos sus reglas descubriendo qué inferencias aceptamos habitualmente [8]; lo mismo ocurre con la epistemología en general. No hay ninguna dificultad en obtener la conmensuración en teología o moral o crítica literaria cuando estas áreas de la cultura son «normales». En algunos periodos, ha sido tan fácil determinar qué críticos tienen una «percepción justa» del valor de un poema como determinar qué experimentadores son capaces de hacer observaciones exactas y mediciones precisas. En otros periodos —por ejemplo, las transiciones entre los «estratos arqueológicos» que distingue Foucault en la reciente historia intelectual de Europa— puede ser igual de difícil saber qué científicos están ofreciendo explicaciones razonables como saber qué pintores están llamados a la inmortalidad.

2. Kuhn y la inconmensurabilidad

En estos últimos años, los debates sobre la posibilidad de la epistemología frente a la hermenéutica han ganado en concreción como consecuencia de la obra de T. S. Kuhn. Su *Structure of Scientific Revolutoins* debía algo a las críticas de Wittgenstein a la epistemología al uso, pero aplicó esas críticas a la opinión generalmente aceptada de una nueva manera. Desde la Ilustración, y especialmente desde Kant, las ciencias físicas se han considerado como paradigma del conocimiento, con el que había de compararse el resto de la cultura. Las lecciones extraídas por Kuhn de la historia de la ciencia indicaban que la controversia dentro de las ciencias físicas se parecía a la conversación ordinaria (sobre la culpabilidad de una acción, los méritos de un candidato, el valor de un poema, la conveniencia de la legislación) más de lo que había dado a entender la Ilustración. En especial, Kuhn se preguntaba si la filosofía de la ciencia podría construir un algoritmo para elegir entre las teorías científicas. Las dudas a este respecto hicieron que sus lectores dudaran doblemente en la cuestión de si la epistemología podría, partiendo de la ciencia, abrirse camino y salir hacia el resto de la cultura descubriendo el

[8] La actitud pragmatista de Nelson Goodman hacia la lógica aparece muy bien resumida en un pasaje que, una vez más, trae a la memoria el «círculo hermenéutico»: «Esto parece un círculo vicioso flagrante... Pero es un círculo virtuoso... *Una regla se corrige si produce una inferencia que no estamos dispuestos a aceptar; una inferencia se rechaza si viola una regla que no estamos dispuestos a corregir*» (*Fact, Fiction and Forecast* [Cambridge, Mass.], 1955, página 67).

terreno común de todo el sector del discurso humano que pudiera considerar como «cognitivo» o «racional».

Los ejemplos de cambio «revolucionario» en la ciencia propuestos por Kuhn eran, como ha comentado él mismo, casos pertenecientes a la clase que la hermenéutica ha considerado siempre como especialmente asignados a ella —casos en que un científico ha dicho cosas que suenan tan absurdas que cuesta creer que le hayamos entendido adecuadamente. Kuhn dice que a sus alumnos les propone esta máxima:

> Al leer las obras de un pensador importante, buscad primero los aparentes absurdos que haya en el texto y preguntaos cómo es posible que una persona sensata los haya escrito. Cuando encontréis una respuesta..., cuando estos pasajes tengan sentido, quizá comprobéis qué pasajes más importantes, aquellos que anteriormente pensabais haber entendido, han cambiado de significado [4].

Kuhn dice que esta máxima no hace falta recordársela a los historiadores, quienes, «consciente o inconscientemente, practican todos el método hermenéutico». Pero la invocación de esta máxima por parte de Kuhn desconcertó a los filósofos de la ciencia que, trabajando dentro de la tradición epistemológica, se veían obligados a pensar en términos de un esquema neutro («lenguaje de la observación», «leyes puente», etc.) que haría conmensarubles, por ejemplo, a Aristóteles y Newton. Este esquema se podía utilizar, pensaban ellos, para hacer innecesarias las conjeturas hermenéuticas. La afirmación de Kuhn de que no hay conmensurabilidad entre grupos de científicos que tienen paradigmas diferentes de una explicación acertada, o que no comparten la misma matriz disciplinar, o ambas cosas [5], a muchos filósofos les produjo la impresión de que ponía en peligro la idea de la elección de teorías en la ciencia. La «filosofía de la ciencia» —nombre con que circulaba la «epistemología» cuando se escondía entre los empiristas lógicos— se había concebido a sí misma como suministradora de un algoritmo para elección de teorías.

La afirmación de Kuhn de que no era posible ningún algoritmo a no ser que fuera *post factum* y Whiggish (que construyera una

[4] T. S. Kuhn, *The Essential Tension* (Chicago, 1977), pág. xii.

[5] Véase «Second Thoughts on Paradigms» (en *ibid.*), sobre la distinción de Kuhn entre dos sentidos «centrales» de «paradigma», mezclados en *The Structure of Scientific Revolutions* pero ahora distinguidos —«paradigma» en cuanto resultado conseguido y paradigma en cuanto «matriz disciplinar».

epistemología basándose en el vocabulario o suposiciones del bando vencedor en una disputa científica) se vio oscurecida, sin embargo, por los addenda de corte «idealista» del propio Kuhn. Una cosa es decir que el «lenguaje de observación neutral» en que los defensores de las diferentes teorías pueden prsentar sus testimonios sirve de poco para optar entre las teorías. Y otra cosa es decir que no puede haber un lenguaje así porque los defensores «ven cosas diferentes» o «viven en mundos diferentes». Por desgracia, Kuhn hizo comentarios incidentales del segundo tipo, y los filósofos se abalanzaron sobre él. Kuhn quería oponerse a la afirmación tradicional de que «lo que cambia con un paradigma es únicamente la interpretación hecha por el científico de observaciones que la naturaleza del entorno y del aparato perceptivo han fijado de una vez por todas» [6]. Pero esta afirmación es inocua si quiere decir meramente que los resultados del mirar pueden formularse siempre en términos aceptables a ambos bandos («el fluido parecía más oscuro», «la aguja giró hacia la derecha», etc.). Kuhn debería haberse contentado con demostrar que la posibilidad de contar con un lenguaje tan inocuo no sirve absolutamente de nada al tener que tomar una decisión entre teorías bajo un algoritmo, lo mismo que no serviría para tomar decisiones sobre la culpabilidad o inocencia en los procesos judiciales bajo un algoritmo, y por las mismas razones. El problema es que el abismo entre el lenguaje neutral y los únicos lenguajes útiles para decidir el problema en cuestión es demasiado grande para salvarlo mediante «postulados de significado» o cualquiera de las otras entidades mitológicas que invocaba la epistemología empirista tradicional.

Kuhn debería haberse limitado a descartar por completo el proyecto epistemológico. Pero, en vez de ello, exigió «un sustituto viable del paradigma epistemológico tradicional» [7] y dijo que «debemos aprender a dar sentido a afirmaciones que al menos se parezcan a 'las obras científicas futuras [después de la revolución] en un mundo futuro'». Pensaba que debíamos dar también sentido a las afirmaciones de que «cuando Aristóteles y Galileo miraban piedras que oscilaban, el primero veía una caída constreñida y el segundo un péndulo» y de que «los péndulos llegaron a existir gracias a algo muy parecido a un interruptor gestáltico inducido por un paradigma». El resultado negativo de estas afirmaciones fue hacer que el péndulo volviera a oscilar entre el realismo y el idealismo. Para

[6] T. S. Kun, *The Structure of Scientific Revolutions*, 2.ª edición (Chicago, 1970), pág. 120.

[7] Ésta y todas las demás citas de Kuhn aprecidas en este párrafo están tomadas de la obra citada, págs. 120-121.

ponernos en guardia ante las confusiones del empirismo tradicional, no debemos pensar qe el interruptor gestáltico en cuestión fuera algo más que el hecho de que las personas llegaran a ser capaces de responder a las estimulaciones sensoriales con observaciones sobre los péndulos, sin tener que hacer una inferencia interviniente. Kuhn tenía razón al decir que había que acabar con «un paradigma filosófico iniciado por Descartes y desarrollado en la misma época de la dinámica newtoniana», pero dejó que su noción de lo que se tenía por «paradigma filosófico» se estableciera con la idea kantiana de que el único sustituto de una explicación realista de un reflejo eficaz era una explicación idealista de la maleabilidad del mundo reflejado. Es cierto que debemos abandonar la idea de «datos e interpretación» con su sugerencia de que si pudiéramos llegar a los datos *reales,* incontaminados por nuestra elección de lenguaje, estaríamos «poniendo las bases» de una opción racional. Pero podemos librarnos de esta idea siendo conductistas en epistemología, más que siendo idealistas. La hermenéutica no necesita un nuevo paradigma epistemológico, lo mismo que el pensamiento político liberal no necesita un nuevo paradigma de la soberanía. La hermenéutica es, más bien, lo que nos queda cuando dejamos de ser epistemológicos.

Dejando de lado el «idealismo» incidental de Kuhn podemos centrarnos simplemente en su afirmación de que no hay ningún algoritmo para la elección entre teorías. Esto ha llevado a sus críticos a decir que con eso permitía que cada científico estableciera su propio paradigma y luego definiera la objetividad y la racionalidad en términos de tal paradigma —crítica que, como he dicho antes, se ha hecho habitualmente a cualquier teoría del conocimiento holística, no fundacional. Por eso Kuhn escribió:

> Al aprender un paradigma, el científico adquiere al mismo tiempo teoría, métodos y criterios, generalmente en una mezcla inextricable...
>
> Esa observación... constituye la primera indicación explícita de por qué la elección entre paradigmas competidores plantea generalmente cuestiones que no se pueden resolver con los criterios de la ciencia normal... Como la cuestión de los criterios encontrados, sólo se puede responder en términos de criterios que quedan totalmente al margen de la ciencia normal, y es este recurso a criterios externos lo que más claramente hace que los debates sobre paradigmas sean revolucionarios [8].

[8] Kuhn, *The Structure of Scientific Revolutions,* 1.ª edición (Chicago, 1961), páginas 108-109.

Y críticos como Scheffler le interpretaban de esta manera:

> ... la evaluación comparativa de paradigmas rivales puede concebirse como un proceso deliberativo que ocurre en un segundo nivel del discurso... regulado, al menos hasta cierto punto, por criterios comunes apropiados a la discusión de segundo orden. El pasaje que acabamos de citar insinúa, sin embargo, que es imposible compartir criterios de segundo orden. Aceptar un paradigma es aceptar no sólo la teoría y los métodos, sino también gobernar las normas de los criterios que sirven para justificar el paradigma frente a sus rivales... Por eso, las diferencias de paradigma se reflejan inevitablemente hacia arriba, en diferencias de criterio en el segundo nivel. De ahí se deduce que cada paradigma es inevitablemente autojustificador, y que los debates sobre paradigmas no pueden tener objetividad: nos vemos llevados nuevamente a conversiones no racionales como descripción final de los cambios de paradigma dentro de la comunidad científica [9].

Es posible, ciertamente, argumentar que «las diferencias de paradigma se reflejan inevitablemente hacia arriba», pero de hecho Kuhn no lo hizo. Se limitó a decir que este reflejo en el meta-discurso hace más difícil resolver las controversias sobre cambios de paradigma que resolver las controversias dentro de la ciencia normal. Hasta aquí estarían de acuerdo críticos como Scheffler; de hecho, como observó Kuhn, «la mayoría de los filósofos de la ciencia... consideran que el tipo de algoritmo que se ha buscado tradicionalmente no es un ideal muy asequible» [10]. La única cuestión real que separa a Kuhn de sus críticos es si la clase de «proceso deliberativo» que se produce en relación con los cambios de paradigma en las ciencias (proceso que, como demuestra Kuhn en *The Copernican Revolution,* puede extenderse todo un siglo) es de diferente naturaleza que el proceso deliberativo que se produce, por ejemplo, en relación con el cambio del *ancien régime* a la democracia burguesa, o de los «Augustos» a los Románticos.

Kuhn dice que los criterios de elección entre teorías (incluso dentro de la ciencia normal, donde no pueden surgir problemas hermenéuticos) «funcionan no como reglas, que determinan la elección, sino como valores, que influyen en ella» (pág. 331). La mayoría de sus críticos llegarían a admitir hasta esto, pero insistirían en que la cuestión crucial es si podemos encontrar una gama de valores

[9] Israel Scheffler, *Science and Subjectivity* (Indianápolis, 1967), pág. 84.
[10] Kuhn, *Essential Tension,* pág. 326.

específicamente *científicos* que afectara a tal elección, en contraposición a «consideraciones ajenas» (el impacto de la ciencia sobre la teología, el futuro de la vida sobre la tierra, y otras por el estilo) que no habría que dejar entrar en el «proceso deliberativo». Kuhn identifica estos criterios como «precisión, coherencia, competencia, simplicidad y productividad» (pág. 332) —enumeración más o menos standard— y podríamos sentirnos tentados a decir que sería «acientífico» permitir que valores que no fueran éstos influyeran en nuestra elección. Pero los intercambios entre la satisfacción de estos diversos criterios dan lugar a un debate racional interminable. Como dice Kuhn:

> Aunque el historiador pueda encontrar siempre hombres —Priestly, por ejemplo— que no fueron razonables al resistirse [a una nueva teoría] durante tanto tiempo como lo hicieron, no encontrará un punto en que la resistencia resulte ilógica o acientífica [11].

Pero, ¿podemos encontrar una forma de decir que las consideraciones presentadas contra la teoría copernicana por el cardenal Belarmino —las descripciones escriturísticas de la estructura de los cielos— *eran* «ilógicas o acientíficas»? [12] Quizá sea éste el punto en que puedan trazarse con más claridad las líneas divisorias entre Kuhn y sus críticos. Gran parte de la idea del siglo XVII de lo que significaba ser «filósofo», y gran parte de la idea de la Ilustración de lo que significaba ser «racional», depende de que Galileo esté absolutamente en lo cierto y la iglesia absolutamente equivocada. Insinuar que en esto hay lugar para el desacuerdo racional —no solamente para un enfrentamiento de buenos y malos entre la razón y la superstición— es poner en peligro la noción misma de «filosofía», pues pone en peligro la idea de encontrar «un método para encontrar la verdad» que considera como paradigmática la mecánica de Galileo y Newton [13]. Todo un complejo de ideas que se refuerzan mutuamente —la filosofía en cuanto disciplina metodológica distinta

[11] Kuhn, *Structure of Scientific Revolutions,* 2.ª edición, pág. 159.

[12] El papel histórico de las complicadas objeciones de Belarmino a las teorías copernicanas aparece descrito en Giorgio de Santillana, *The Crime of Galileo* (Chicago, 1955). La significación de la postura de Belarmino la estudia Michael Polanyi en *Personal Knowledge* (Chicago, 1958).

[13] Los fundadores de la «filosofía moderna» consideraron paradigmática la mecánica en un doble sentido. Por una parte, (el método de encontrar la verdad» debía ser el que había seguido Newton, o al menos uno que llegara a los resultados newtonianos. Por otra parte, en escritores como Locke, la mecánica newtoniana era un modelo para la mecánica del «espacio interno» (las operaciones mentales «para-mecánicas» satirizadas por Reid y Ryle).

de las ciencias, la epistemología en cuanto fuente de conmensuración, la racionalidad en cuanto posible únicamente basándose en el terreno común que hace posible la conmensuración— parece correr riesgo si se contesta negativamente a la pregunta sobre Belarmino.

Kuhn no ofrece una respuesta explícita a la pregunta, pero en sus escritos se encuentra un arsenal de argumentos en favor de una respuesta negativa. En cualquier caso, la respuesta negativa queda implícita en la argumentación del presente libro. La consideración fundamental es si sabemos trazar una línea divisoria entre ciencia y teología de tal manera que entender bien los cielos sea un valor «científico», y mantener a la Iglesia y la estructura cultural general de Europa sea un valor «acientífico» [14]. El argumento de que no ocurre eso se centra en torno a la afirmación de que las líneas divisorias entre disciplinas, objetos materiales, partes de la cultura, se ven en situación de peligro ante sugerencias novedosas e importantes. Este argumento puede formularse en términos del ámbito del criterio de la «competencia» —uno de los desiderata clásicos para las teorías, como hemos enumerado antes. Belarmino pensaba que la competencia de la teoría de Copérnico era más reducida de lo que pudiera pensarse. Cuando sugería que la teoría copernicana quizá no fuera realmente más que un ingenioso dispositivo heurístico en orden a la navegación, por ejemplo, y otros tipos de cálculos celestes de orientación práctica, estaba admitiendo que la teoría, con sus propios límites, era exacta, coherente, simple y quizá hasta productiva. Cuando decía que no debería pensarse que su competencia fuera mayor de lo indicado, defendía su actitud diciendo que tenían excelentes testimonios independientes (de la Escritura) para creer que los cielos eran a grandes rasgos tolemaicos. ¿Procedían sus testimonios de otra esfera y, por tanto, la restricción que propone era «acientífica»? ¿Qué es lo que determina que la Escritura no sea una excelente fuente de comprobación de cómo están organizados los cielos? Montones de cosas, especialmente la decisión de la Ilustración

---

[14] Otro ejemplo del mismo tipo es la cuestión suscitada en torno a la «objetividad» por los críticos marxistas de las distinciones tradicionales entre las áreas de la cultura. Véase, por ejemplo, Herbert Marcuse, *One-Dimensional Man* (Boston, 1964), capítulos 6-7. Más en concreto, podemos preguntar si hay una forma clara de separar el valor «científico» de conseguir la heredabilidad del derecho a la inteligencia del valor «político» de reducir el racismo. Creo que Marcuse tiene razón cuando dice que para hacer esta distinción hace falta la mayor parte del aparato intelectual («burgués») de la Ilustración. Sin embargo, a diferencia de Marcuse, yo tengo la esperanza de que podríamos mantener la distinción incluso después de rechazar una pieza del aparato —la filosofía «fundacional» epistemológicamente orientada.

de que la Cristiandad era en su mayor parte fruto del clericalismo. Pero, ¿qué se suponía que debían decir los contemporáneos de Belarmino —que creían en su mayor parte que la Escritura era realmente la palabra de Dios— a Belarmino? Lo que dijeron, entre otras cosas, fue que la fidelidad a la Escritura podía desvincularse de la admisión de ciertas ideas adventicias (por ejemplo, aristotélicas y tolemaicas) que se habían utilizado para interpretar la Escritura. (Es lo mismo que dirían los teólogos liberales del siglo XIX a propósito del Génesis y de Darwin). ¿Estaban mal orientados todos estos argumentos sobre hasta qué punto podría ser legítimamente liberal la interpretación que cada uno hacía de la Biblia? Eran argumentos orientados a limitar, por así decirlo, el ámbito de competencia de la Escritura (y, por tanto, de la iglesia) —la reacción contraria a la del propio Belarmino de limitar el ámbito de competencia de Copérnico. Por eso, la cuestión sobre si Belarmino (y, necesariamente, los defensores de Galileo) estaba introduciendo consideraciones «acientíficas» ajenas parece ser la pregunta de si hay alguna forma antecedente de determinar la relevancia de una afirmación para otra, alguna «red» (por utilizar el término de Foucault) que determine qué clases de testimonio *podría* haber para las afirmaciones sobre los movimientos de los planetas.

Evidentemente, la conclusión que quiero sacar es que la «red» que apareció a finales del siglo XVII y en el XVIII era algo a lo que se podía acudir a comienzos del siglo XVII, en la época en que fue procesado Galileo. Antes de su aparición, no había epistemología ni estudio de la naturaleza del conocimiento humano que pudiera haberla «descubierto». La idea de qué era ser «científico» estaba en proceso de formación. Si se aceptan los valores —o, quizá la clasificación de los valores competidores— comunes a Galileo y a Kant, entonces es cierto que Belarmino estaba siendo «acientífico». Pero, por supuesto, la mayoría de nosotros (incluyendo a Kuhn, aunque quizá no a Feyerabend) los aceptamos con gusto. Somos herederos de trescientos años de retórica sobre la importancia de distinguir claramente entre ciencia y religión, ciencia y política, ciencia y arte, ciencia y filosofía, etc. Esta retórica ha formado la cultura de Europa. Nos ha hecho ser lo que somos hoy en día. Tenemos la ventaja de que la existencia de una pequeña perplejidad dentro de la epistemología o dentro de la historiografía de la ciencia es suficiente para acabar con ella. Pero proclamar nuestra lealtad a estas distinciones no es decir que haya normas «objetivas» y «racionales» para su adopción. Galileo, podríamos decir, impuso su argumento, y todos nosotros estamos en el terreno común de la «red» de relevancia e

irrelevancia que desarrolló la «filosofía moderna» como consecuencia de esa victoria. Pero, ¿qué podría demostrar que el problema Belarmino-Galileo «sea de naturaleza diferente» al problema que hay, por ejemplo, entre Kerensky y Lenin, o entre la Royal Academy (en torno al año 1910) y Bloomsbury?

Puedo explicar la idea de «diferencia de naturaleza» que está en cuestión aquí remitiendo a la noción de conmensurabilidad. La diferencia deseada es aquella que nos permitiría decir que cualquier observador razonable desinteresado de la controversia Belarmino-Galileo, teniendo en cuenta todas las consideraciones relevantes, vendría a ponerse de parte de Galileo, mientras que hombres razonables todavía difieren en las otras cuestiones que acabo de mencionar. Pero esto nos hace volver otra vez a la cuestión de si los valores que invocaba Belarmino eran propiamente «científicos», si su actitud pasa por «desinteresada» y sus pruebas por «relevantes». En este punto, me parece a mí, haríamos bien en abandonar la idea de ciertos valores («racionalidad», «desinterés») que se mantienen al margen de los patrones educativos e institucionales de la época. Podemos decir simplemente que Galileo estaba *creando* la idea de «valores científicos sobre la marcha, que fue una gran cosa que lo hiciera, y que está fuera de lugar la cuestión de si era racional al hacerlo.

Como dice Kuhn en relación con un problema menor, pero evidentemente relacionado con éste, no podemos diferenciar a las comunidades científicas por «el objeto material», sino más bien «examinando los patrones de educación y comunicación» [15]. Saber qué es lo que cuenta como relevante para una elección entre teorías sobre una cierta materia es, en periodos de investigación normal, pertenecer a lo que Kuhn llama «matriz disciplinar». En periodos en que está en cuestión la comunidad relevante de investigadores, en que se están borrando las líneas divisorias entre «eruditos», «meros empíricos» y chapuceros (o, por cambiar el ejemplo, entre «pensadores políticos serios» y «panfletistas revolucionarios»), la cuestión de la relevancia es libre. No podemos determinar la relevancia concentrando la atención en el objeto material y diciendo, por ejemplo: «No os preocupéis por lo que diga la Escritura que hizo Dios; basta con mirar los planetas y ver qué hacen *ellos*». El *mero* mirar a los planetas no servirá para nada a la hora de elegir nuestro modelo de los cielos, como tampoco el *mero* leer la Escritura. En 1550, había un determinado conjunto de consideraciones que eran relevantes para las concepciones «racionales» en astronomía, y al llegar el año 1750

[15] Kuhn, *Essential Tension,* pág. xvi.

había un conjunto de consideraciones relevantes que era en gran parte diferente. Este cambio en lo que se consideraba relevante puede entenderse, retrospectivamente, como el trazado de distinciones adecuadas entre lo que había realmente en el mundo («descubrir» que la astronomía era una esfera autónoma de la investigación científica), o como un cambio en el clima cultural. No importa mucho que lo consideremos de una manera o de otra, con tal que tengamos claro que el cambio no fue producido por un «argumento racional» en un sentido de «racional» en que, por ejemplo, los cambios introducidos últimamente en relación con la actitud de la sociedad hacia la esclavitud, el arte abstracto, los homosexuales o las especies de animales en peligro, *no* figuraran como «racionales».

Resumiendo la línea que vengo siguiendo a propósito de Kuhn y sus críticos: la controversia entre ellos sobre si la ciencia, en cuanto descubrimiento de lo que está ahí realmente en el mundo, difiere en sus patrones de argumentación de los discursos donde parece menos adecuada la idea de «correspondencia con la realidad» (por ejemplo, la política y la crítica literaria). La filosofía lógico-empirista de la ciencia, y toda la tradición epistemológica a partir de Descartes, ha tratado de decir que el procedimiento para conseguir representaciones precisas en el Espejo de la Naturaleza difiere en formas importantes del procedimiento para conseguir acuerdos sobre cuestiones «prácticas» o «estéticas». Kuhn nos da razones para decir que no hay diferencia más profunda que la que se da entre lo que ocurre en el discurso «normal» y el «anormal». Esta distinción atenta contra la distinción entre ciencia y no ciencia. La terrible indignación con que fue recibida la obra de Kuhn [16] era natural, pues los ideales de la Ilustración no son sólo nuestro patrimonio cultural más precioso sino que están en peligro de desaparición en la medida en que los estados totalitarios van extendiéndose a una parte cada vez mayor de la humanidad. Pero el hecho de que la Ilustración uniera el ideal de la autonomía de la ciencia ante la teología y la política con la imagen de la teoría científica como Espejo de la Naturaleza no es

[16] No obstante, la virulencia se manifestó principalmente entre los filósofos profesionales. La descripción de Kuhn de cómo funciona la ciencia no sorprendió a los científicos cuya racionalidad trataban de proteger los filósofos. Pero los filósofos unían a una vinculación profesional a la metáfora del espejo una comprensión del papel fundamental que habían desempeñado estas metáforas en la Ilustración y, por tanto, en hacer posible la base institucional de la ciencia moderna. Tenían razón al ver que la crítica de Kuhn a la tradición era profunda, y que la ideología que había protegido la ascensión de la ciencia moderna estaba en peligro. Se equivocaban al pensar que las instituciones seguían necesitando la ideología.

razón para mantener esta confusión. La red de relevancia e irrelevancia que heredamos casi intacta desde el siglo XVIII será más atractiva cuando ya no esté unida a esta imagen. Las desfasadas imágenes del espejo no sirven de nada para mantener intacta la herencia —moral y científica— de Galileo.

3. La objetividad como correspondencia y como acuerdo

Los críticos de Kuhn han contribuido a perpetuar el dogma de que sólo donde hay correspondencia con la realidad donde hay posibilidad de acuerdo racional, en un sentido especial de «racional» cuyo paradigma es la ciencia. Esta confusión se ve fomentada por nuestro uso de «objetivo» en sentido de «que caracteriza la concepción en que estaríamos de acuerdo como consecuencia de un argumento no perturbado por consideraciones irrelevantes» y de «que representa las cosas tal como son». Los dos sentidos son en gran parte coextensivos, y fuera del contexto filosófico no supondría ningún problema unirlos. Pero si comenzamos a tomarnos en serio preguntas como «¿En qué sentido exactamente está la Bondad ahí fuera esperando a ser representada exactamente como consecuencia de una argumentación racional sobre cuestiones morales?» o «¿En qué sentido estaban ahí los rasgos físicos de la realidad, capaces de ser representados exactamente sólo mediante ecuaciones diferenciales, o tensores, antes de que las personas pensaran en representarlos de esa manera?», comienza a notarse la tensión entre las dos ideas. Debemos agradecer a Platón la pregunta del primer tipo y al idealismo y pragmatismo la del segundo. Ninguna de las dos preguntas tiene respuesta. Nuestra inclinación natural a responder con un rotundo «En ningún sentido» a la primera y un igualmente rotundo «En el más pleno de los sentidos posibles y en el más directo» a la segunda no nos servirá de nada para liberarnos de estas preguntas si seguimos sintiendo la necesidad de *justificar* las respuestas a estas preguntas construyendo teorías epistemológicas y metafísicas.

Desde Kant, el principal empleo de estas teorías ha sido apoyar intuiciones relacionadas con la distinción subjetivo-objetivo —o intentos de demostrar que nada exterior a la ciencia natural cuenta como «objetivo» o intentos de aplicar este término honorífico a la moral, o a la política o a poemas. La metafísica, como intento de averiguar qué es aquello sobre lo que se puede ser objetivo, se ve obligada a preguntar por las semejanzas y diferencias que hay, por ejemplo, en el descubrimiento (como consecuencia de la resolución

final de un dilema moral planteado desde hacía mucho tiempo) de un nuevo artículo de la Ley Moral, el descubrimiento (por los matemáticos) de un nuevo tipo de número o de un nuevo conjunto de espacios, el descubrimiento de la indeterminación del quántum, y el descubrimiento de que el gato está en la alfombra. Este último descubrimiento —un *point d'appui* casero para las nociones de «contacto con la realidad», «la verdad como correspondencia» y «exactitud de la representación»— es la medida con que se comparan los otros en cuanto a objetividad. Por eso, el metafísico ha de preocuparse por los sentidos en que los valores, números y paquetes de ondas se parecen a los gatos. El epistemólogo ha de preocuparse por los sentidos en que afirmaciones más interesantes comparten la objetividad poseída por ese triunfo del reflejo —la elocución adecuada de «El gato está en la alfombra». Según la concepción que se desprende del conductismo epistemológico, no hay ninguna manera interesante de descubrir si, por ejemplo, hay una Ley Moral con la que pueda establecerse una correspondencia. El hecho de que, por ejemplo, «las normas morales implicadas por la naturaleza del hombre» encajen mejor en el universo hilemórfico de Aristóteles que en el mecanicista de Newton, no es razón para pensar que haya o no haya una Ley Moral «objetiva». Ni podría serlo ninguna otra cosa. El problema de la metafísica, tal como decían los positivistas, es que nadie tiene claro qué debe considerarse como un argumento satisfactorio dentro de ella, aunque, por supuesto, lo mismo ocurre en el caso de la filosofía «impura» del lenguaje que practicaron los positivistas (por ejemplo, la tesis de Quine de la «no-factualidad» de lo intencional). Según el punto de vista que estoy defendiendo, podríamos, en una época imaginaria en que el consenso dentro de estas áreas fuera completo, considerar la moralidad, la física y la psicología igualmente «objetivas». Entonces podríamos relegar las áreas más discutibles de la crítica literaria, la química y la sociología al campo de lo «no-cognitivo» o «interpretarlas operacionalísticamente» o «reducirlas» a una u otra disciplina «objetiva». La aplicación de títulos honoríficos como «objetivo» y «cognitivo» no es nunca más que una expresión de la presencia, y la esperanza, de un acuerdo entre los investigadores.

Anque quizá suponga repetir algo de lo que hemos dicho anteriormente, creo que vale la pena volver a considerar el debate entre Kuhn y sus críticos, ahora en el contexto de una discusión de la distinción «objetivo-subjetivo», sencillamente por el interés de esta distinción y por su carga de sentido moral. Una vez más, este sentido moral es consecuencia de la idea (totalmente justificada) de que

nuestra mejor esperanza es la conservación de los valores de la Ilustración. Por eso, en esta sección intentaré una vez más cortar los lazos que unen estos valores con la imagen del Espejo de la Naturaleza.

Es conveniente comenzar con la forma en que el propio Kuhn se enfrenta a la afirmación de que su concepción es una puerta abierta a la «subjetividad». Dice:

> «Subjetivo» es un término con varios usos establecidos: en uno de ellos se opone a «objetivo», en otro a «basado en un juicio». Cuando mis críticos describen los rasgos idiosincrásicos a que yo recurro diciendo que son subjetivos, utilizan, y yo creo que erróneamente, el segundo de estos sentidos. Cuando se quejan de que dejo a la ciencia sin objetividad, confunden el segundo sentido de subjetivo con el primero [17].

En el sentido en que los «rasgos subjetivos» no hacen referencia a un juicio, son, continúa Kuhn, «cuestión de gusto» —el tipo de cuestión que nadie se preocupa en discutir, simples informes del propio estado mental. Pero, naturalmente, el valor de un poema o de una persona no es, en este sentido, cuestión de gusto. Por eso, puede decir Kuhn, el valor de una teoría científica es cuestión de «juicio, no de gusto» en el mismo sentido.

Esta respuesta a la acusación de «subjetividad» es útil dentro de su campo de acción, pero no llega a afectar al temor más profundo que hay detrás de la acusación. Es el temor a que no haya realmente ningún terreno intermedio entre las cuestiones de gusto y las cuestiones que se pueden solucionar con un algoritmo previamente formulable. El filósofo que no ve este terreno intermedio está argumentando más o menos, pienso yo, como sigue:

1. Todas las afirmaciones describen o estados internos de los seres humanos (su Esencia de Vidrio, el Espejo quizá empañado) o estados de la realidad exterior (naturaleza).
2. Podemos saber de qué clase es cada afirmación viendo sobre cuáles sabemos conseguir un acuerdo universal.
3. Por eso, la posibilidad de desacuerdo perpetuo es índice de que, independientemente de lo racional que parezca ser el debate, no hay realmente nada sobre lo que debatir —pues el tema sólo pueden ser estados internos.

Este curso de razonamiento, compartido por los platónicos y los po-

[17] Kuhn, *Essential Tension,* pág. 336.

sitivistas, produjo en los últimos la noción de que «analizando» oraciones podemos descubrir si tratan de hecho sobre lo «subjetivo» o lo «objetivo» —donde «análisis significa averiguar si hay acuerdo general entre los hombres sensatos y racionales sobre lo que debería contar como confirmación de su verdad. Dentro de la epistemología tradicional, sólo ocasionalmente se ha visto esta última idea tal como es: la admisión de que nuestra única noción utilizable de «objetividad» es «acuerdo», más que reflejo. Incluso en la gratificante y franca observación de Ayer de que «definimos una creencia racional como aquella a la que se llega por métodos que ahora consideramos fiables» [18], la idea de «fiabilidad» todavía sigue funcionando como una insinuación de que sólo podemos ser racionales por correspondencia con lo real. Ni siquiera su reconocimiento igualmente franco de que todas las representaciones privilegiadas del mundo le permitirán a un hombre «mantener sus convicciones frente a pruebas aparentemente hostiles si está dispuesto a hacer las necesarias suposiciones *ad hoc*» (pág. 95) basta para acabar con la convicción de Ayer de que al separar lo «empírico» de lo «emotivo» y lo «analítico» está separando la «verdad sobre el mundo» de algo diferente. Esto se debe a que Ayer, como Platón, añade a la anterior cadena de razonamiento otra premisa fundacionalista:

4. Sólo somos capaces de eliminar la posibilidad de desacuerdo racional perpetuo, irresoluble, en aquellas áreas donde los vínculos incuestionados con la realidad externa proporcionan un terreno común a los que disputan.

La afirmación de que, donde no podemos encontrar vínculos incuestionados con (por ejemplo, representaciones privilegiadas de) los objetos que se deben reflejar, no hay posibilidad de que haya un algoritmo, junto con la afirmación de que donde no hay posibilidad de que haya un algoritmo sólo puede haber la *apariencia* de acuerdo racional, lleva a la conclusión de que la ausencia de representaciones privilegiadas relevantes demuestra que estamos únicamente ante «una cuestión de gusto». Kuhn tiene razón cuando dice que esto es muy distinto de la idea común de «gusto», pero, como la noción igualmente poco corriente de verdad en cuanto algo que no necesita tener nada que ver con el acuerdo, tiene una larga historia dentro de la filosofía [19]. Es una historia que hay que entender para ver

[18] A. J. Ayer, *Language, Truth and Logic* (Nueva York, 1970), pág. 100.

[19] Kant introdujo esta noción de la verdad sobre la realidad en la filosofía alemana (y, *a fortiori*, en la filosofía en cuanto disciplina profesionalizada

por qué unas sugerencias historiográficas tan poco fuertes como las de Kuhn removieron los niveles inconscientes más profundos de la mente filosófica adiestrada.

Quizá la mejor forma de hacer frente a la acusación de «subjetivismo» formulada contra Kuhn es establecer una distinción entre los sentidos de «subjetivo» distinta de la que él mismo ha hecho en el pasaje citado. Podemos distinguir dos sentidos de «subjetivo» que ocupan oposiciones casi contrarias a cada uno de los dos sentidos de «objetivo» distinguidos anteriormente. La «objetividad», en el primer sentido, era una propiedad de las teorías que, tras haber sido ampliamente discutidas, son elegidas por un consenso de los que las han discutido de forma racional. En contraste, una consideración «subjetiva» es aquella que ha sido dejada de lado, o lo sería, o debería de serlo, por las personas que han participado en la discusión de forma racional —la que se considera, o debería considerarse, que es irrelevante para el objeto material de la teoría. Decir que alguien está introduciendo consideraciones «subjetivas» en una discusión donde se exige objetividad es, en términos generales, decir que está introduciendo consideraciones que otros consideran que están fuera de lugar. Si insiste en estas consideraciones *outré,* está convirtiendo la investigación normal en un discurso anormal. Que una consideración sea subjetiva, en este sentido, es sencillamente que sea poco conocida. Por eso, juzgar la subjetividad es tan arriesgado como juzgar la relevancia.

En un sentido más tradicional, «subjetivo» contrasta con «que corresponde a lo que está ahí fuera», y, por tanto, significa algo parecido a «producto únicamente de lo que hay aquí dentro» (en el corazón, o en la «confusa» porción de la mente que no contiene representaciones privilegiadas y, por tanto, no refleja exactamente lo que hay ahí fuera). En este sentido, «subjetivo» va asociado a «emocional» o «fantástico», pues nuestros corazones y nuestras imaginaciones son idiosincrásicos, mientras que nuestros entendimientos son, en sus mejores momentos, espejo idéntico de los objetos externos mismos. Aquí tenemos una conexión con las «cuestiones

que tenía su modelo en las universidades alemanas). Lo hizo distinguiendo el mero hacer frente a los fenómenos y el intuir intelectualmente los noúmenos. También introdujo en la cultura europea la distinción entre juicio cognitivo y estético, y la distinción entre este último y el mero gusto. No obstante, dentro de los límites de la actual discusión, su distinción entre «juicio estético» que puede ser acertado o equivocado y «gusto» que no puede serlo, quedará marginada. Los críticos de Kuhn podrían criticarle con más cautela (pero, según sus puntos de vista, con la misma fuerza devastadora) por hacer de la elección de teoría en la ciencia una cuestión de juicio estético más que cognitivo.

de gusto», pues el estado de nuestras emociones en un momento dado (del que tenemos un ejemplo en nuestra reacción momentánea no deliberada ante una obra de arte) es algo que no se presta a discusión. Tenemos acceso privilegiado a lo que ocurre dentro de nosotros. De esta manera, la tradición procedente desde Platón ha unido la distinción «algoritmo versus no algoritmo» con la distinción «razón versus pasión». Las diversas ambigüedades de «objetivo» y «subjetivo» ilustran cómo puede desarrollarse la confusión. Si no fuera por la vinculación tradicional de estas distinciones, un historiador de la investigación que insistiera en las semejanzas entre las controversias de los científicos con las de los críticos literarios no habría sido interpretado como si estuviera poniendo en peligro nuestras mentes al elevar de categoría a nuestros corazones.

Sin embargo, el propio Kuhn hace ocasionalmente concesiones demasiado amplias a la tradición, especialmente cuando señala que hay un problema serio y sin solucionar en torno a las razones por las que la empresa científica ha marchado tan bien últimamente. Así, dice:

> Incluso los que me han seguido hasta aquí querrán saber cómo es posible que una empresa basada en el valor, como la que yo he descrito, pueda desarrollarse tal como lo hace la ciencia, produciendo repetidamente nuevas técnicas poderosas para la predicción y control. Por desgracia, no tengo respuesta ninguna para esta pregunta, pero eso no es más que otra forma distinta de decir que no pretendo haber resuelto el problema de la inducción. Si la ciencia progresara en virtud de algún algoritmo de elección compartido y vinculante, estaría igualmente desorientado a la hora de explicar sus éxitos. Se trata de una laguna que siento intensamente, pero su presencia no diferencia mi postura de la de la tradición[20].

[20] Kuhn, *Essential Tension,* págs. 332-333. Hay otros pasajes del libro donde, debo reconocer, Kuhn es demasiado condescendiente con la tradición epistemológica. Uno está en la pág. xxiii, donde expresa la esperanza de que la comprensión por los filósofos de «determinación de referencia y traducción» ayude a clarificar los problemas. Otro está en la pág. 14, donde dice que la filosofía de la ciencia tiene una misión totalmente distinta de las actividades hermenéuticas del historiador de la ciencia: «La filosofía se ocupa de la reconstrucción racional, y sólo debe preservar aquellos elementos de su materia que sean esenciales para la ciencia en cuando conocimiento razonado.» Me parece que estas líneas conservan intacto el mito de que hay algo que se llamaba «la naturaleza del conocimiento razonado» que los filósofos deben describir, actividad totalmente distinta de la descripción de lo que se tiene por justificación dentro de las diversas matrices disciplinares que constituyen la cultura del momento.

Como he dicho al hablar del «realismo metafísico» de Putnam, en el capítulo anterior, *no* hace falta sentir tanto esa laguna. No hemos de lamentarnos de nuestra incapacidad de conseguir una hazaña que nadie tiene la más mínima idea de cómo realizar. La idea de que nos enfrentamos con el reto de llenar esta laguna es un resultado más de la hipostatización del *focus imaginarius* platónico —la verdad en cuanto separada del acuerdo— y de permitir que el abismo entre uno mismo y esa idea incondicional nos haga sentir que no llegamos a entender las condiciones de la propia existencia.

Según lo que yo vengo proponiendo, la pregunta «¿Por qué, si la ciencia es meramente..., produce nuevas técnicas poderosas de predicción y control?» es como la pregunta «¿Por qué, si el cambio en la conciencia moral de Occidente desde 1750 es meramente..., ha podido conseguir tanto en favor de la libertad humana?» Podemos rellenar el primer espacio en blanco con «adhesión al siguiente algoritmo vinculante...» o con «una sucesión de matrices disciplinarias institucionalizadas kuhnianas». El segundo lo podemos completar con «la aplicación del pensamiento secular a las cuestiones morales» o «la conciencia culpable de la burguesía» o «cambios en la constitución emocional de los que controlan las claves del poder» o con muchas otras expresiones. En ninguno de los casos sabe nadie lo que podría ser una buena respuesta. Con visión retrospectiva, «Whiggish» y «realista», siempre podremos ver el logro deseado (predicción y control de la naturaleza, emancipación de los oprimidos) como resultado de tener una visión más clara de lo que hay ahí (los electrones, las galaxias, la Ley Moral, los derechos humanos). Pero éstas no son nunca las explicaciones que buscan los filósofos. Son, en fórmula de Putnam, explicaciones «internas» —explicaciones que satisfacen nuestra necesidad de contar un relato causal y coherente sobre nuestras interacciones con el mundo, pero no nuestra necesidad trascendental de confirmar nuestro reflejo haciendo ver cómo se aproxima a la verdad. «Resolver el problema de la inducción», en el sentido a que se refiere Kuhn, sería como «resolver el problema del hecho y el valor»; ambos problemas sobreviven únicamente como nombres de una cierta insatisfacción inarticulada. Son el tipo de problemas que no se pueden formular dentro de la «filosofía normal»; lo único que ocurre es que algún que otro artilugio técnico recibe ocasionalmente el nombre de «solución» a tal problema, con la vaga esperanza de establecer contacto con el pasado, o con la eternidad.

Lo que necesitamos, más que una solución al «problema de la inducción», es la capacidad de pensar sobre la ciencia de tal manera

que el que sea una «empresa basada en el valor» no nos ocasione ninguna sorpresa. Lo único que nos lo impide es la idea firmemente arraigada de que los «valores» son «internos» mientras que los «hechos» son «externos» y que resulta igual de misterioso cómo, empezando con los valores, podemos producir bombas, que cómo, empezando con episodios internos privados, podemos evitar darnos golpes con las cosas. Con esto volvemos al enojoso problema del «idealismo» y a la idea de que la búsqueda de un algoritmo va unida a un enfoque «realista» de la ciencia, mientras que el dejarse llevar por el método meramente hermenéutico del historiador es pasarse al idealismo. Siempre que se sugiere que se mitiguen las distinciones entre teoría y práctica, hecho y valor, método y conversación, se sospecha que hay un intento de hacer al mundo «maleable a la voluntad humana». Esto provoca, a su vez, la afirmación positivista de que debemos hacer una distinción clara entre lo «no-cognitivo» y lo «cognitivo» o bien «reducir» lo primero a lo segundo. La tercera posibilidad —reducir lo segundo a lo primero— parece que «espiritualiza» la naturaleza al hacerla igual a la historia o la literatura, es decir, algo que los hombres han *hecho* y no algo que se *encuentran*. A algunos críticos les parece que Kuhn está proponiendo la tercera opción.

Este intento renovado de ver a Kuhn como si tendiera al «idealismo» no es más que una forma confusa de reiterar la afirmación de que algo como (4) es verdadero —que debemos considerar a los científicos «en contacto con la realidad externa» y, por tanto, con posibilidades de llegar a un acuerdo racional por medios de que no disponen los políticos ni los poetas. La confusión está en sugerir que Kuhn, al «reducir» los métodos de los científicos a los de los políticos, ha «reducido» el mundo «encontrado» de los neutrones al mundo «hecho» de las relaciones sociales. Una vez más nos encontramos con la idea de que todo lo que no puede ser descubierto por una máquina programada con el algoritmo adecuado no puede existir «objetivamente» y, por tanto, ha de ser de alguna manera una «creación humana». En la sección siguiente, trataré de unir lo que vengo diciendo sobre la objetividad con algunos temas procedentes de las páginas anteriores del libro, con la esperanza de hacer ver que la distinción entre epistemología y hermenéutica no debe considerarse como paralela a la distinción entre lo que está «ahí fuera» y lo que nosotros «inventamos».

4. Espíritu y naturaleza

Hay que confesar que la idea de que hay un conjunto especial de métodos adecuados para las disciplinas «blandas» —las *Geisteswissenschaften*— tiene vínculos históricos con el idealismo. Como dice Apel, la actual oposición entre filosofía analítica y «hermenéutica» como estrategias filosóficas parece natural desde que

> la metafísica del espíritu y del sujeto del Idealismo del siglo XIX, que debían ser consideradas como fundamentos de las *Geisteswissenschaften* (aunque las últimas ponían más énfasis, ciertamente, en la investigación material) son interpretadas por el Wittgenstein tardío como una «enfermedad» del lenguaje junto con otros conceptos de la metafísica de la filosofía occidental[21].

La idea de que podría entregarse el yo empírico a las ciencias de la naturaleza, pero no el yo trascendental, que constituye el mundo fenoménico y (quizá) funciona como agente moral, ha contribuido tanto como lo que más a dar significado a la distinción espíritu-naturaleza. Por eso, esta distinción metafísica está oculta en el trasfondo de toda discusión de las relaciones entre *Geistes-* y *Naturwissenschaften*. La imagen se complica todavía más con la noción vaga de que los que tienen afición a hablar de «hermenéutica» están proponiendo introducir un nuevo tipo de método (sospechosamente «blando») en vez de algún otro método (el «método científico», por ejemplo, o quizá el «análisis filosófico»). En esta sección espero demostrar que la hermenéutica, en cuanto discurso sobre discursos todavía-inconmensurables, no tiene ninguna conexión especial ni con (a) el lado de la «mente» del dualismo cartesiano, ni con (b) el lado «constituyente» de la distinción kantiana entre la facultad constituyente» de la distinción kantiana entre la facultad constituyente y estructurante de la espontaneidad y la facultad pasiva de la receptividad, ni (c) con la idea de un método que permita descubrir la verdad de las oraciones que compita con los métodos normales seguidos en las disciplinas extrafilosóficas. (Sin embargo, creo que este sentido limitado y purificado de «hermenéutica» que estoy utilizando *sí* está ligado al uso del término por escritores como Gadamer, Apel y Habermas. Trataré de explicitar las conexiones en el siguiente capítulo.)

---

[21] Karl-Otto Apel, *Analytic Philosophy of Language and the Gesteswissenschaften* (Dordrecht, 1967), pág. 35. Cfr., también, pág. 53.

El miedo a «caer en el idealismo» que afecta a quienes se han visto tentados por Kuhn a rechazar las ideas consagradas de la filosofía de la ciencia (y, más en general, de la epistemología) se incrementa con la idea de que si el estudio de la búsqueda científica de la verdad respecto al universo físico se interpreta hermenéuticamente se interpretará como la actividad del espíritu —la facultad que *hace*— y no como aplicación de las facultades reflectoras, las que *encuentran* lo que la naturaleza ha hecho ya. Esta oposición latente romántico-clásico que se mantiene oculta en el trasfondo del examen de Kuhn se manifiesta abiertamente en el uso poco afortunado (condenado en la sección 2, más arriba) que hace Kuhn de expresiones románticas de «encontrarse con un mundo nuevo», en vez del clásico «uso de una nueva descripción para el mundo». Según el planteamiento que estoy siguiendo aquí, no hay nada importante que dependa de la elección entre estas dos expresiones —entre las imágenes de hacer y encontrar. Es una oposición parecida a la que se da entre «objetivo» y «no objetivo» o «cognitivo» y «no cognitivo», de que he hablado en la sección anterior. Sin embargo, resulta menos paradójico mantenerse fiel a la noción clásica de «describir mejor lo que ya estaba allí» en el caso de la física. Esto no se debe a consideraciones epistemológicas o metafísicas profundas, sino simplemente a que, cuando contamos nuestros relatos «Whiggish» sobre cómo nuestros antepasados escalaron gradualmente la montaña en cuya cumbre (quizá falsa) estamos ahora, debemos mantener algunas cosas constantes a lo largo del relato. Las fuerzas de la naturaleza y los pequeños fragmentos de la materia, tal como los entiende la actual teoría física, son buenos ejemplos de este papel. La física es el paradigma de «encontrar» sencillamente porque es difícil (al menos en el Occidente) contar una historia de universos físicos cambiantes con el fondo de una Ley Moral o canon poético invariables, y en cambio resulta muy fácil contar la historia contraria. Nuestro obstinado sentido «naturalista» de que el espíritu es, si no reducible a la naturaleza, al menos dependiente de ella, no es más que la idea de que la física nos ofrece un trasfondo adecuado sobre el cual colocar nuestras historias del cambio histórico. No es como si tuviéramos una visión profunda de la naturaleza de la realidad que nos dijera que todo, menos los átomos y el vacío, era «por convención» (o «espiritual» o «inventado»). La idea de Demócrito era que un relato sobre los fragmentos más diminutos de las cosas constituye un buen trasfondo para las historias sobre los cambios de las cosas hechas de estos fragmentos. La aceptación de este género de historia del mundo (enriquecida sucesivamente por Lucrecio,

Newton y Bohr) quizá sea un rasgo definitorio de Occidente, pero no es una opción que pueda conseguir, o que exija, garantías epistemológicas o metafísicas.

Yo concluyo que los seguidores de Kuhn deberían resistir a la tentación de ganar la partida a los Whigs hablando de «mundos diferentes». Renunciando a estas expresiones, no concederían nada a la tradición epistemológica. Decir que el estudio de la historia de la ciencia, como el estudio del resto de la historia, debe ser hermenéutico, y negar (como haría yo, pero no Kuhn) que haya algo extra llamado «reconstrucción racional» que pueda *legitimar* la práctica científica actual, no es decir todavía que los átomos, paquetes de ondas, etc., descubiertos por los científicos físicos sean creaciones del espíritu humano. Aprovecharse de la ciencia de la época en la elaboración del relato más amplio posible sobre la historia de la raza no es, a no ser que se adquieran también los diversos dogmas platónicos mencionados en la sección anterior, decir que la física sea «objetiva» en algún sentido en que quizá no lo sean la política o la poesía. La línea que separa el hacer del encontrar no tiene nada que ver con la línea divisoria entre inconmensurabilidad y comensurabilidad. O, dicho de otra manera, el sentido en que un hombre es espiritual y no meramente un ser natural —el sentido a que han prestado atención antirreduccionistas como Alasdair MacIntyre, Charles Taylor y Marjorie Grene— no es un sentido en que sea un ser que haga el mundo. Decir, con Sartre, que el hombre se hace a sí mismo, y que por eso mismo difiere de los átomos y de los tinteros, es perfactamente compatible con el rechazo de toda insinuación de que parte de su auto-creación consista en «constituir» átomos y tinteros. Por las confusiones entre la noción romántica del hombre como auto-creativo, la idea kantiana del hombre en cuanto que constituye un mundo fenoménico y la idea cartesiana del hombre en cuanto que contiene un ingrediente inmaterial especial merecen un estudio más detallado. Este conjunto de confusiones aparece materializado en gran parte de la discusión referida a la «naturaleza del espíritu», «la irreductibilidad de la persona», la distinción entre acción y movimiento, y la distinción entre *Geistes* y *Naturwissenschaften*. Como la última distinción es probablemente coextensiva con la distinción entre métodos hermenéuticos y no hermenéuticos, es especialmente importante considerarla para clarificar la noción de hermenéutica que estoy presentando.

Para comenzar a desenmarañar esta triple confusión examinaré la afirmación de que la hermenéutica se acomoda especialmente al «espíritu» o a «las ciencias del hombre», mientras hay algún otro mé-

todo (el de las ciencias «objetivizantes» y «positivas») que es apropiado para la «naturaleza». Si trazamos la línea divisoria entre epistemología y hermenéutica tal como vengo haciendo —como contraste entre el discurso sobre el discurso normal y el anormal— parece claro que no compiten entre sí, sino más bien que se ayudan mutuamente. Nada hay tan valioso para el investigador hermenéutico de una cultura exótica como el descubrimiento de una epistemología escrita dentro de esa cultura. Nada es tan valioso para la determinación de si los posesores de esa cultura han expresado alguna verdad interesante (según —¿qué otra cosa podría ser?— los criterios del discurso normal de nuestro propio tiempo y lugar) como el descubrimiento hermenéutico de cómo traducirlas sin hacer que parezcan unos locos. Por eso, sospecho que esta noción de métodos competidores procede de la idea de que el mundo aparece dividido en áreas que reciben su mejor descripción en el discurso normal (el «esquema conceptual», por utilizar una expresión pre-davidsoniana) de nuestra propia cultura, y otras áreas en que no ocurre eso. En concreto, esta concepción sugiere que las personas van a ser siempre de alguna manera tan viscosas y resbaladizas (la «viscosidad» de Sartre) que escaparán a la explicación «objetiva». Pero, una vez más, si se traza la distinción hermenéutica-epistemología tal como yo lo hago, no hay ninguna necesidad de que sea más difícil entender a las personas que a las cosas; lo único que ocurre es que la hermenéutica sólo se necesita en el caso de discursos inconmensurables. Como señalan acertadamente los fisicistas, en el momento en que podemos saber traducir lo que se está diciendo, no hay ninguna razón para pensar que la explicación de por qué se está diciendo deba diferir en naturaleza, o seguir métodos diferentes) de una explicación de la locomoción o la digestión. No hay ninguna razón metafísica por la que los seres humanos deban ser capaces de decir cosas inconmensurables, ni ninguna garantía de que vayan a seguir haciéndolo. Es sólo cuestión de buena suerte (desde un punto de vista hermenéutico) o de mala suerte (desde un punto de vista epistemológico) que lo hayan hecho así en el pasado.

La disputa tradicional sobre la «filosofía de las ciencias sociales» se ha desarrollado en líneas generales de la siguiente manera. Un bando ha dicho que la «explicación» (subsumir bajo leyes predictivas, más o menos) presupone, y no puede reemplazar, a la «comprensión». El otro bando ha dicho que la comprensión *es* simplemente la capacidad de explicar, que lo que sus oponentes llaman «comprensión» no es más que la etapa primitiva de ir buscando a

tientas algunas hipótesis explicativas. Ambos bandos tienen razón. Apel comenta, acertadamente, que

> los protagonistas de la «comprensión» (es decir, de las *Geisteswissenschaften)* atacan a los defensores de la teoría de la explicación (es decir, de las ciencias conductuales o sociales objetivas) desde atrás —y viceversa—. Los «científicos objetivos» señalan que los resultados de la «comprensión» sólo tienen validez precientífica y subjetivamente heurística, y que deben, cuando menos, ser comprobados y complementados con métodos analíticos objetivos. Los protagonistas de la comprensión, por el contrario, insisten en que la obtención de datos en las ciencias sociales —y, por tanto, cualquier comprobación objetiva de las hipótesis— presupone la «comprensión real»... del significado [22].

Los que desconfían de la hermenéutica quieren decir que el hecho de que algunos seres hablen no es ninguna razón para pensar que escapan a la gran red unificada de las poderosas leyes de predicción, pues estas leyes pueden prever lo que van a decir así como lo que van a comer. Los que defienden la hermenéutica dicen que la cuestión de lo que van a decir se divide en dos partes —los sonidos o inscripciones que realizan (que podrían ser bastante previsibles, quizá gracias a la neurofisiología), y lo que *significan,* que es algo muy distinto. En este punto, la actitud natural en los defensores de la «ciencia unificada» es decir que *no* es diferente, pues hay procedimientos para traducir cualquier evolución significativa a una lengua concreta —el lenguaje de la misma ciencia unificada. Dado un lenguaje único que contiene todo lo que puede decir cualquiera (tal como Carnap intentaba ensamblar ese lenguaje en la *Aufbau),* la pregunta de qué oración de lenguaje está siendo propuesta por el usuario del lenguaje sometido a investigación no es más «especial» que la pregunta de qué va a tomar para cenar. La traducción al lenguaje de la ciencia unificada es difícil, pero el intento de traducir no implica técnicas de construcción o comprobación de teorías distintas del intento de explicar los hábitos alimenticios.

En respuesta a esto, los defensores de la hermenéutica deberían limitarse a decir que, más como cuestión de hecho innegable que de necesidad metafísica, no existe eso que llaman el «lenguaje de la ciencia unificada». No *tenemos* un lenguaje que sirva como matriz neutra y permanente para formular todas las buenas hipótesis explicativas, y no tenemos la menor idea de cómo conseguir uno.

---

[22] *Ibíd.,* pág. 30.

(Esto es compatible con decir que *sí* tenemos un lenguaje de observación que es neutro, aunque inútil.) Por eso no es probable que la epistemología —en cuanto intento de hacer conmensurables todos los discursos traduciéndolos a un conjunto de términos preferido— sea una estrategia útil. La razón no es que la «ciencia unificada» sólo funcione en un ámbito metafísico y no en otro, sino que la suposición «Whiggish» de que tengamos ese lenguaje bloquea el camino de la investigación. Siempre podríamos desear cambiar el lenguaje hablado por los sujetos de nuestra explicación. Pero esto sería sólo un caso especial de la posibilidad permanente de que alguien tenga una idea mejor. La comprensión del lenguaje hablado por los sujetos, el entender las explicaciones que dan de por qué están haciendo esto o aquello, puede ser útil o no. En el caso de las personas que son especialmente estúpidas, o psicópatas, hacemos bien en dejar de lado sus explicaciones. Les atribuimos intenciones y acciones en términos que no aceptan y quizá ni entiendan. La conocida afirmación de que para determinar la acción que está realizando un sujeto hay que tener generalmente en cuenta la descripción de sí mismo hecha por el hablante, es bastante acertada. Pero es perfectamente posible dejar de lado dicha descripción. El privilegio que se le atribuye es moral, más que epistémico. La diferencia entre su descripción y la nuestra puede significar, por ejemplo, que debería ser juzgado con nuestras leyes. No significa que no pueda ser explicado por nuestra ciencia.

Decir que no podemos entender una cultura extraña si nos empeñamos en interpretarla como si tuviera «demasiados» de nuestros deseos y creencias es sólo una generalización de la observación kuhniana de que no podemos entender a los científicos del pasado si nos obstinamos en hacer lo mismo con ellos. Esto puede generalizarse a su vez en la afirmación de que no debemos suponer que el vocabulario utilizado hasta ahora valdrá para todo lo demás que vaya apareciendo. El problema no es que los espíritus se resistan por naturaleza a ser objetos de previsión, sino simplemente que no hay ninguna razón para pensar (y muchas razones para no pensar) que nuestro propio espíritu esté ahora en posesión del mejor de los vocabularios para formular hipótesis que expliquen y prevean a todos los demás espíritus (o, quizá, a los demás cuerpos). Esto es lo que dice Charles Taylor, quien formula la cuestión como sigue:

> ... podríamos asustarnos tanto ante la perspectiva de esta ciencia hermenéutica que deseemos volver al modelo verificacionista. ¿Por qué no podemos considerar nuestra comprensión del sig-

> nificado como parte de la lógica del descubrimiento, tal como sugieren los empiristas lógicos para nuestras ideas no formalizables, y a pesar de ello basar nuestra ciencia en la exactitud de nuestras predicciones? [23]

y responde enumerando tres razones por las que «esta predicción exacta es radicalmente imposible». Dice que

> la tercera razón, y la más importante, de la imposibilidad de predicción rigurosa es que el hombre es un animal que se autodefine. Con los cambios en su autodefinición vienen los cambios en lo que es el hombre, de forma que ha de ser entendido en términos diferentes. Pero las mutaciones conceptuales de la historia humana pueden producir, y producen frecuentemente, redes conceptuales que son inconmensurables, es decir, en las que los términos no se pueden definir en relación con un estrato común de expresiones (pág. 49).

La afirmación de que lo que dificulta la predicción de la conducta de los habitantes de la cultura extraña es simplemente la inconmensurabilidad de su lenguaje me parece totalmente correcta, pero creo que luego Taylor oscurece su observación cuando continúa diciendo:

> El éxito de la predicción en las ciencias naturales está estrechamente ligado con el hecho de que todos los estados del sistema, del pasado y del futuro, se pueden describir dentro del mismo ámbito de conceptos, como valores, por ejemplo, de las mismas variables. Por eso, todos los estados futuros del sistema solar se pueden describir dentro del mismo ámbito de conceptos, como valores, por ejemplo, de las mismas variables. Por eso, todos los estados futuros del sistema solar se pueden describir, lo mismo que los pasados, en el lenguaje de la mecánica newtoniana... Sólo si el pasado y el futuro se meten en la misma red es posible entender los estados de lo posterior como función de los estados de lo anterior y, por tanto, hacer una previsión.
>
> Esta unidad conceptual queda viciada en las ciencias del hombre por el hecho de la innovación conceptual que a su vez altera la realidad humana (pág. 49).

Taylor restablece aquí la idea del hombre como ser que cambia desde dentro encontrando formas mejores (o, al menos, nuevas) de describir, predecir y explicarse a sí mismo. Los seres humanos, como meros *êtres-en-soi,* no se cambian desde dentro, simplemente son descritos, previstos y explicados en un vocabulario mejor. Esta forma

[23] Charles Taylor, «Interpretation and the Sciences of Man», *Review of Metaphysics,* 25 (1971), pág. 38.

de decirlo nos lleva de nuevo a la desafortunada idea metafísica antigua de que el universo está hecho de dos clases de cosas. El sentido en que los seres humanos se cambian a sí mimos redescribiéndose no es metafísicamente más interesante o misterioso que el sentido en que se cambian variando de dieta, pareja sexual o residencia. Se trata del mismo sentido: a saber, resultan verdaderas en su caso oraciones nuevas y más interesantes. Taylor continúa diciendo que «los mismos términos en que deberá describirse el futuro para entenderlo adecuadamente no están actualmente a nuestra disposición» (pág. 50), y pretende que esto sólo ocurre en el caso de los seres humanos. Pero, por lo que podemos saber, puede ser que se haya agotado la creatividad humana, y que en el futuro sea el mundo *no*-humano el que se escape de nuestra red conceptual. Podría ocurrir que todas las futuras sociedades humanas sean (como consecuencia, quizá, de un totalitarismo tecnocrático omnipresente) vulgares variaciones del nuestro. Pero la ciencia contemporánea (que parece ya incapaz de explicar la acupuntura, la migración de las mariposas, etc.) quizá llegue pronto a encontrarse en una situación tan mala como el hilemorfismo de Aristóteles. La línea que está describiendo Taylor no es la línea entre lo humano y lo no humano, sino entre esa parte de la investigación donde no estamos seguros de disponer del vocabulario adecuado y esa parte donde estamos bastante seguros de poseerlo. Por el momento, coincide, a grandes rasgos, con la distinción entre los campos de las *Geistes-* y *Naturwissenschaften*. Pero esta coincidencia puede que sea una *mera* coincidencia. Con una perspectiva suficientemente larga, puede resultar que el hombre sea menos δεινός de lo que pensaba Sófocles, y las fuerzas elementales de la naturaleza lo sean más de lo que se imaginan los fisicistas modernos.

Para ver mejor este punto, conviene tener presente que hay muchas ocasiones en que ignoramos por completo el *pour-soi* de los seres humanos. Lo hacemos en el caso de personas particularmente grises y convencionales, por ejemplo, cuyas acciones y palabras son todas tan previsibles que los «objetivizamos» sin dudarlo. Por el contrario, cuando tropezamos con algo no humano que trata de escaparse de la red conceptual utilizada en el momento, es natural que se comience a hablar de un lenguaje desconocido —imaginar, por ejemplo, que las mariposas migratorias tienen un lenguaje en que describen rasgos del mundo para los que la mecánica newoniana no tiene nombre. O, si no llegamos tan lejos, al menos incurrimos, naturalmente, en la idea de que todavía no se ha conseguido descifrar el Libro de la Naturaleza —que no contiene, por ejemplo, la

«gravitación» más de lo que contiene el «movimiento natural». La tentación es antropomorfizar el mundo no humano, o parte de él, en el momento en que resulte claro que, como ocurre con el nativo de una cultura exótica, o el genio cuya conversación nos supera, no «hablamos el mismo lenguaje». Naturaleza es todo aquello que es tan de rutina y tan familiar y tan dócil que nos permite confiar implícitamente en nuestro propio lenguaje. Espíritu es todo aquello que es tan extraño e incontrolable que nos hace empezar a preguntarnos si nuestro «lenguaje» es «adecuado» para ello. Nuestra duda, despojada de las imágenes del espejo, se refiere simplemente a si no es posible que alguien o algo esté relacionándose con el mundo en términos para los que nuestro lenguaje no tiene equivalentes. Dicho más sencillamente, sólo es preguntarse si no necesitamos cambiar nuestro vocabulario, y no solamente nuestras afirmaciones.

Al comenzar este capítulo he dicho que la hermenéutica es, a grandes rasgos, una descripción de nuestro estudio de lo poco conocido, y la epistemología es, a grandes rasgos, la descripción de nuestro estudio de lo que nos resulta familiar. Dada la interpretación algo forzada que acabo de proponer de «espíritu» y «naturaleza», podría estar ahora de acuerdo con la opinión tradicional de que la hermenéutica describe nuestra investigación del espíritu, mientras que la epistemología es una descripción de nuestra investigación de la naturaleza. Pero sería mejor, pienso yo, abandonar por completo la distinción espíritu-naturaleza. Como ya he dicho antes, esta distinción une *a)* la distinción entre lo que encaja perfectamente en nuestra forma de explicar y predecir las cosas y lo que no; *b)* la distinción entre algo que une todas las distintas características (enumeradas en el capítulo primero, sección 3) que se han considerado, en una u otra época, como distintivamente humanas, y el resto del mundo. Además une estas dos distinciones con *c)* la distinción (criticada en el capítulo tercero, sección 3) entre la facultad de la espontaneidad (la actividad transcendental de la constitución) y la de la receptividad. Lo hace por una combinación de nuestra facultad receptiva transcendental del sentido con el campo de las presentaciones sensoriales que integran el «yo empírico» (combinación que el propio Kant fue incapaz de evitar). El resultado de unir el espíritu como creatividad romántica auto-transcendente (siempre dispuesto a comenzar a hablar en una forma que sea inconmensurable con nuestro actual lenguaje) con el espíritu en cuanto idéntico a la Esencia de Vidrio del hombre (con toda su libertad metafísica de la explicación física), y con el espíritu como «constituyente» de la realidad fenoménica, fue la metafísica del idealismo alemán del

siglo XIX. Fue un conjunto de asimilaciones provechoso, pero uno de sus resultados menos afortunados fue la idea de que la filosofía tenía una esfera especial y propia, totalmente alejada de la de la ciencia. Esta asimilación contribuyó a mantener viva la idea de la «filosofía» como disciplina centrada en la epistemología. Mientras la idea del espíritu en cuanto constituyente transcendental (en sentido kantiano) estuviera reforzada por el recurso al dualismo cartesiano, por una parte, y al romanticismo, por otra, podría sobrevivir sin problemas la idea de una disciplina que presidiría sobre las demás y se llamaría «epistemología» o «filosofía transcendental» —que no se podría reducir ni a *Naturwissenschaft* (psicofisiología) ni a *Geisteswissenschaft* (la sociología del conocimiento). Otro legado poco feliz fue la confusión de la necesidad de traducción no mecánica (y, más en general, de formación de conceptos imaginativa) con la «irreductibilidad del ego transcendental constituyente». Esta confusión mantuvo viva la discusión idealismo-realismo mucho después de cuando debería haberse dado por concluida, pues los amigos de la hermenéutica pensaron (como ilustra la cita de Apel al comienzo de esta sección) que algo parecido al idealismo era la carta oficial de su actividad, mientras que sus enemigos supieron que cualquiera que practicara manifiestamente la hermenéutica tenía que ser «antinaturalista» y carecer de un sentido adecuado de la exterioridad bruta del universo físico.

Para resumir lo que quiero decir sobre la «irreductibilidad de las *Geisteswissenschaften*», voy a proponer las siguientes tesis:

El fisicismo tiene probablemente razón cuando dice que algún día seremos capaces, «en principio», de predecir todos y cada uno de los movimientos del cuerpo de una persona (incluyendo los de su laringe y de la mano con que escribe) por referencia a microestructuras existentes dentro de su cuerpo.

El peligro para la libertad humana como consecuencia de este logro es mínimo, pues las palabras «en principio» admiten la probabilidad de que la determinación de las condiciones iniciales (los estados antecedentes de las microestructuras) sea demasiado difícil de conseguir, a no ser que se trate de algún ejercicio pedagógico ocasional. En cualquier caso, los que se dedican a las torturas y al lavado de cerebros se encuentran ya en buena posición para entorpecer la libertad humana según se les antoje; el posterior progreso científico no puede mejorar su posición.

La intuición existente detrás de la distinción tradicional entre naturaleza y espíritu, y detrás del romanticismo, es que podemos predecir qué sonidos saldrán de la boca de alguien sin saber lo

que significan. Así pues, aun cuando pudiéramos predecir los sonidos emitidos por la comunidad de los investigadores científicos del año 4.000, no estaríamos por eso en condiciones de participar en su conversación. Esta intuición es totalmente correcta [24].

El hecho de que podamos predecir un sonido sin saber lo que significa es precisamente el hecho de que las condiciones microestructurales necesarias y suficientes para la producción de un sonido en muy pocas ocasiones correrán parejas a una equivalencia material entre una afirmación en el lenguaje utilizado para describir la microestructura y la afirmación expresada por el sonido. Esto no se debe a que algo sea en principio imprevisible, y mucho menos a una división ontológica entre naturaleza y espíritu, sino simplemente a la diferencia entre un lenguaje adecuado para enfrentarse con las neuronas y otro adecuado para enfrentarse con las personas.

Podemos saber cómo responder a una observación críptica de un juego lingüístico diferente sin saber ni preocuparnos de qué oración de nuestro juego lingüístico corriente es materialmente equivalente a esa observación [25]. Producir la conmensurabilidad encontrando equivalencias materiales entre oraciones procedentes de distintos juegos lingüísticos es sólo una de las distintas técnicas existentes para hacer frente a nuestros semejantes. Cuando no funciona, recurrimos a otra que funcione —por ejemplo, coger el truco a un nuevo juego lingüístico, y quizá olvidar el antiguo. Ésta es la misma técnica que utilizamos cuando la naturaleza no humana se manifiesta recalcitrante a dejarse predecir en el vocabulario de la ciencia tradicional.

La hermenéutica no es «otra forma de conocer» —la «com-

---

[24] Éste es el núcleo de verdad oculto en la afirmación quineana de que las *Geisteswissenschaften* no contienen «cuestiones de hecho». Raymond Geuss lo expresa bien hablando de Quine: «Aun cuando tengamos una teoría de la naturaleza que nos permita predecir las disposiciones verbales de alguien para toda la eternidad, con ello no entenderemos lo que *quiere decir*» («Quine und die Unbestimmtheit der Ontologie», *Neue Hefte für Philosophie,* Heft 8 [1973], pág. 44n.).

[25] El que todas las lenguas sean traducibles unas a otras (por las razones davidsonianas mencionadas en el capítulo 6) no significa que se puedan encontrar tales equivalencias (al menos «en principio»). Significa simplemente que no podemos dar sentido a la afirmación de que nuestros conocimientos tienen algo más que impedimentos temporales —la afirmación de que una cosa que se llama «esquema conceptual diferente» nos impide aprender a conversar con el usuario de otro lenguaje. Tampoco elimina la intuición que había tras la falsa afirmación romántica de que los grandes poemas son intraducibles. Son traducibles, por supuesto; el problema es que las traducciones no son ya grandes poemas.

prensión» en oposición a la «explicación» (predictiva). Es mejor considerarla como otra forma de arreglárselas. Sería mejor para la claridad filosófica que entregáramos la idea de «cognición» a la ciencia predictiva, y dejáramos de preocuparnos por los «métodos cognitivos alternativos». La palabra *conocimiento* no parecería digna de que se luchara por ella si no fuera por la tradición kantiana de que ser filósofo es tener una «teoría del conocimiento», y la tradición platónica de que la acción que no está basada en el conocimiento de la verdad de las proposiciones es «irracional».

Capítulo VIII

Filosofía sin espejos

1. Hermenéutica y edificación

Las ideas actuales de lo que significa ser un filósofo están tan vinculadas con el intento kantiano de hacer conmensurables todas las pretensiones de conocimiento que es difícil imaginar qué podría ser la filosofía sin la epistemología. En un plano más general, es difícil imaginar que ninguna actividad se considerara digna de llevar el nombre de «filosofía» si no tuviera nada que ver con el conocimiento —si no fuera en cierto sentido una teoría del conocimiento, un método de conseguir conocimiento, o al menos una pista de dónde podría encontrarse una clase de conocimiento de la máxima importancia. La dificultad procede de una idea que es común a los platónicos, kantianos y positivistas: que el hombre tiene una esencia —a saber, descubrir esencias. La idea de que nuestra tarea principal es reflejar con exactitud, en nuestra propia Esencia de Vidrio, el universo que nos rodea, es el complemento a la idea, común a Demócrito y Descartes, de que el universo está formado por cosas muy simples, clara y distintamente cognoscibles, el conocimiento de cuyas esencias constituye el vocabulario-maestro que permite la conmensuración de todos los discursos.

Hay que dejar de lado esta imagen clásica de los seres humanos antes de poder dejar de lado una filosofía cuyo centro esté en la epistemología. El intento de conseguirlo tiene el nombre de «hermenéutica», término polémico dentro de la filosofía contemporánea. El uso del término en este contexto se debe en gran parte a una obra —*Truth and Method*, de Gadamer—. En ella, Gadamer deja claro que la hermenéutica no es un «método para conseguir la verdad» que encaje en la imagen clásica del hombre: «El fenómeno hermenéutico

no es básicamente un problema de método, en absoluto» [1]. Más bien lo que hace Gadamer, a grandes rasgos, es preguntar qué conclusiones se pueden deducir del hecho de que tenemos que practicar la hermenéutica —del «fenómeno hermenéutico» en cuanto un hecho que la tradición epistemológica ha tratado de marginar. «La hermenéutica que se desarrolla aquí», dice, no es... una metodología de las ciencias humanas, sino un intento de entender qué son verdaderamente las ciencias humanas, más allá de su autoconciencia metodológica, y qué las conecta con la totalidad de nuestra experiencia del mundo» [2]. Su libro es una nueva descripción del hombre que intenta colocar la imagen clásica dentro de otra más amplia y, por tanto, de «distanciar» la problemática filosófica consagrada en vez de ofrecer un conjunto de soluciones a la misma.

En relación con el tema que nos ocupa, la importancia de la obra de Gadamer está en que consigue separar una de las tres hebras —la idea romántica del hombre en cuanto auto-creador— existentes en la idea filosófica del «espíritu» de las otras dos hebras con que se había llegado a confundir. Gadamer (como Heidegger, con quien está en deuda parte de su obra) no hace concesiones ni al dualismo cartesiano ni a la noción de «constitución trascendental» (en cualquier sentido que pudiera recibir una interpretación idealista) [3]. De esta manera contribuye a reconciliar la afirmación «naturalista» que he intentado realizar en el capítulo precedente —que la «irreductibilidad de las *Geisteswissenschaften*» no es cuestión de dualismo metafísico— con nuestra intuición «existencialista» de que lo más importante que podemos hacer es redescribirnos a nosotros mismos. Lo hace utilizando la noción de Bildung (educación, auto-formación) en vez de la de «conocimiento», en cuanto meta del pensamiento. Decir que nos convertimos en otras personas, que nos «rehacemos» a nosotros mismos al leer más, hablar más y escribir más, no es más que una forma llamativa de decir que las oraciones que resultan

---

[1] Hans-Georg Gadamer, *Truth and Method* (Nueva York, 1975), pág. xi. En realidad, podría decirse que el libro de Gadamer es un panfleto contra la misma idea de método, donde éste se concibe como un intento de conmensuración. Resulta instructivo observar los paralelismos entre este libro y el de Paul Feyerabend, *Against Method*. En la exposición de Gadamer sigo a Alasdair MacIntyre; véase su «Contexts of Interpretation», *Boston University Journal*, 24 (1976), 41-46.

[2] Gadamer, *Truth and Method*, pág. xiii.

[3] Cfr. *ibid.*, pág. 15. «Pero podemos reconocer que la *Bildung* es un elemento del espíritu sin sentirnos ligados a la filosofía hegeliana del espíritu absoluto, lo mismo que la idea de la historicidad de la conciencia no está unida a su filosofía de la historia del mundo.»

verdaderas de nosotros en virtud de tales actividades son, con frecuencia, mucho más importantes para nosotros que las oraciones que resultan verdaderas de nosotros cuando bebemos más, ganamos más, etcétera. Los hechos que nos hacen capaces de decir cosas nuevas e interesantes sobre nosotros mismos son, en este sentido, no-metafísico, más «esenciales» para nosotros (al menos para nosotros, intelectuales de vida relativamente tranquila, habitantes de una parte próspera y estable del mundo) que los hechos que cambian nuestras formas o nuestros niveles de vida («re-hacernos» en formas menos «espirituales»). Gadamer desarrolla su idea de *wirkungsgeschichtliches Bewusstsein* (el tipo de conciencia del pasado que nos cambia) para describir una actitud interesada, no tanto en lo que hay ahí fuera en el mundo, o en lo que ocurrió en la historia, cuanto en lo que podemos sacar de la naturaleza y de la historia para nuestros propios usos. Cuando se tiene esta actitud, la obtención de hechos correctos (sobre los átomos y el vacío, o sobre la historia de Europa) sólo tiene valor propedéutico para encontrar una forma nueva y más interesante de expresarnos a nosotros mismos y, por tanto, de arreglárnoslas con el mundo. Desde el punto de vista educacional, en oposición al epistemológico o tecnológico, la forma en que se dicen las cosas es más importante que la posesión de verdades[4].

Como la palabra «educación» suena demasiado vulgar, y la de *Bildung* nos resulta extraña, utilizaré el término «edificación» para referirme a este proyecto de encontrar nuevas formas de hablar que sean más interesantes y provechosas. El intento de edificar (a nosotros mismos y a los demás) puede consistir en la actividad hermenéutica de establecer conexiones entre nuestra propia cultura y alguna cultura o periodo histórico exóticos, o entre nuestra propia disciplina y otra disciplina que parezca buscar metas inconmensurables con un vocabulario inconmensurable. Pero puede consistir también en la actividad «poética» de elaborar esas metas nuevas,

---

[4] El contraste que se da en este caso es el mismo que el que se aprecia en la discusión tradicional entre educación «clásica» y educación «científica» mencionado por Gadamer en la sección inicial de «The Significance of the Humanist Tradition». Más en general, puede considerarse como un aspecto de la lucha entre poesía (que no puede abandonarse en la primera clase de educación) y filosofía (que, cuando se concibe a sí misma como una superciencia, desearía convertirse en fundamento de la segunda clase de educación). Yeats preguntó a los espíritus (que, creía él, le estaban dictando *A Vision* por mediación de su esposa) por qué habían venido. Los espíritus respondieron: «Para traeros metáforas para la poesía.» Un filósofo habría esperado algunos datos más concretos sobre lo que ocurría al otro lado, pero Yeats no se sintió defraudado.

nuevas palabras, o nuevas disciplinas, a lo que seguiría, por así decirlo, lo contrario de la hermenéutica: el intento de reinterpretar nuestros entornos familiares en términos, no familiares, de nuestras nuevas invenciones. En cualquier caso, la actividad es (a pesar de la relación etimológica entre las dos palabras) edificar sin ser constructivo —al menos si «constructivo» significa aquella forma de cooperación en la realización de los programas de investigación que tiene lugar en el discurso normal. Se *supone* que el discurso que edifica es anormal, que nos saca de nosotros mismos por la fuerza de lo extraño, para ayudarnos a convertirnos en seres nuevos.

El contraste entre deseo de edificación y deseo de verdad no es, según Gadamer, expresión de una tensión que exija su resolución o la adopción de un compromiso. Si hay conflicto, éste se da entre la concepción platónico-aristotélica de que la *única* forma de ser edificado es saber qué hay ahí fuera (reflejar los hechos con exactitud —realizar nuestra esencia conociendo esencias) y la concepción de que la búsqueda de la verdad es sólo una de las muchas formas en que podemos ser edificados. Gadamer atribuye justamente a Heidgger el mérito de haber dado con una forma de ver la búsqueda de conocimiento objetivo (desarrollada primeramente por los griegos, utilizando como modelo las matemáticas) como uno de entre los diversos proyectos humanos [5]. Sin embargo, éste queda reflejado más gráficamente en Sartre, quien concibe el intento de adquirir un conocimiento objetivo del mundo y, por tanto, de uno mismo, como un intento de evitar la responsabilidad de elegir el propio proyecto [6]. Para Sartre, decir esto no es decir que el deseo de conocimiento objetivo de la naturaleza, de la historia o de cualquier otra cosa esté condenado al fracaso, ni siquiera que tenga que resultar engañoso. Es decir, simplemente, que presenta una tentación al auto-engaño en la medida en que pensemos que, sabiendo qué descripciones dentro de un conjunto dado de discursos normales se aplican a nosotros, por eso mismo nos conocemos a nosotros. Para Heidegger, Sartre y Gadamer, la investigación objetiva es perfectamente posible y muchas veces real —lo único que hay que decir en su contra es que proporciona sólo algunas, de las mu-

---

[5] Véase la sección titulada «The Overcoming of the Epistemological Problem...», en *Truth and Method,* págs. 214 y ss., y compárese con Martin Heidegger, *Being and Time,* traducido por John Macquarrie y Edward Robinson (Nueva York, 1962), sec. 32.

[6] Véase Jean-Paul Sartre, *Being and Nothingness,* traducción de Hazel Barnes (Nueva York, 1956), parte segunda, capítulo 3, sección 5, y la «Conclusión» del libro.

chas, formas de describirnos a nosotros mismos, y que algunas de ellas pueden dificultar el proceso de edificación.

Resumiendo esta concepción «existencialista» de la objetividad, podríamos decir: la objetividad debe entenderse como conformidad con las normas de justificación (para las afirmaciones y para las acciones) que encontramos sobre nosotros. Esta conformidad sólo resulta dudosa y engañosa cuando se entiende como algo más que esto —a saber, como una forma de obtener acceso a algo que «sirva de base» a las actuales prácticas de justificación en alguna otra cosa. De dicha «base» se piensa que no necesita justificación, pues se ha percibido tan clara y distintamente como para servir de «fundamento filosófico». Esto es engañoso no simplemente por el absurdo general de apoyar la justificación última en lo injustificable, sino también por el absurdo más concreto de pensar que el vocabulario utilizado por la ciencia o la moralidad actual tengan una vinculación privilegiada con la realidad que haga de él algo *más* que otro conjunto cualquiera de descripciones. El acuerdo con los naturalistas en que la redescripción no es «cambio de esencia» debe ir seguido por el abandono total de la noción de «esencia» [7]. Pero la estrategia filosófica consagrada en la mayoría de los naturalismos es encontrar una forma de demostrar que nuestra propia cultura ha conseguido captar la esencia del hombre —convirtiendo así a todos los vocabularios nuevos e inconmensurables meramente en una ornamentación «no-cognitiva» [8]. La utilidad de la opinión «existencialista» es que, al proclamar que no tenemos esencia, nos permite ver las descripciones de nosotros mismos que encontramos en una de (o en la unidad de) las *Naturwissenschaften* en pie de igualdad con las diversas descripciones alternativas presentadas por poetas, novelistas, psicólogos profundos, escultores, antropólogos y místicos. Las primeras no son representaciones privilegiadas en virtud del hecho de que

[7] Habría resultado oportuno que Sartre hubiera completado esta observación de que el hombre es el ser cuya esencia es no tener esencia diciendo que esto valía también para todos los demás seres. Si no se introduce esa adición, parecerá que Sartre está insistiendo en la antigua distinción metafísica entre espíritu y naturaleza, aunque en otros términos, en vez de afirmar simplemente que el hombre es siempre libre de elegir nuevas descripciones (entre otras cosas, para sí mismo).

[8] Creo que Dewey es el único de los autores habitualmente clasificados como «naturalistas» que no ha manifestado esta actitud reductora, a pesar de que habla continuamente del «método científico». El mérito singular de Dewey fue haber permanecido lo suficientemente hegeliano como para no pensar que la ciencia natural tenga un acceso interior a las esencias de las cosas, al mismo tiempo que se mantenía lo suficientemente naturalista como para pensar en los seres humanos en términos darwinianos.

(por el momento) haya más consenso en las ciencias que en las artes. Están simplemente dentro del repertorio de auto-descripciones a nuestra disposición.

El tema puede formularse también como una extrapolación del tópico de que no puede decirse de alguien que esté educado —*gebildet*— si *sólo* conoce los resultados de las *Naturwissenschaften* de la época. Gadamer comienza *Truth and Method* examinando el papel de la tradición humanista en la atribución de sentido a la idea de *Bildung* como algo que «no tiene fines fuera de sí mismo»[9]. Para dar sentido a esta idea hace falta un sentido de la relatividad de los vocabularios descriptivos para los periodos, tradiciones y accidentes históricos. Esto es lo que hace la tradición humanista dentro de la educación, y no puede conseguir el adiestramiento en los resultados de las ciencias naturales. Dado ese sentido de relatividad, no podemos tomarnos en serio la noción de «esencia», ni la idea de que la tarea del hombre sea la representación exacta de las esencias. Las ciencias naturales, por sí solas, no llegan a convencernos de que sabemos lo que somos y lo que podemos saber —no sólo cómo predecir y controlar nuestra conducta, sino los límites de tal conducta (y, en especial, los límites de nuestra habla significativa). El intento de Gadamer de rechazar la exigencia de «objetividad» (en la que coinciden Mill y Carnap) en las *Geisteswissenschaften* es el intento de impedir que la educación se reduzca a una instrucción en los resultados de la investigación normal. Más en general, es el intento de impedir que la investigación anormal sea considerada sospechosa únicamente por su anormalidad.

A este intento «existencialista» por colocar la objetividad, la racionalidad y la investigación normal dentro del cuadro más amplio de nuestra necesidad de ser educados y edificados se enfrenta muchas veces el intento «positivista» de distinguir entre aprender hechos y adquirir valores. Desde el punto de vista positivista, la exposición que hace Gadamer de la *wirkungsgeschichtliche Bewusstsein* puede parecer poco más que la repetición del tópico de que aun cuando sepamos todas las descripciones objetivamente verdaderas de nosotros mismos, quizá sigamos sin saber qué hacer con nosotros. Desde este punto de vista, *Truth and Method* (y los capítulos sexto y séptimo del presente libro) son adaptaciones exageradas del hecho de que la total satisfacción de todas las exigencias de justificación que ofrece la investigación normal nos dejaría en libertad para extraer nuestras propias conclusiones de las afirmaciones así justifi-

[9] Gadamer, *Truth and Method,* pág. 12.

cadas. Pero desde los puntos de vista de Gadamer, Heidegger y Sartre, el problema de la distinción hecho-valor es que está concebida precisamente para oscurecer el hecho de que son posibles otras descripciones además de las presentadas por los resultados de las investigaciones normales [10]. Indica que una vez «están dentro todos los hechos» no queda nada, excepto la adopción «no-cognitiva» de una actitud —opción que no se puede discutir racionalmente. Oculta el hecho de que usar un conjunto de oraciones verdaderas para describirnos a nosotros mismos es ya elegir una actitud hacia nosotros, mientras que utilizar otro conjunto de oraciones verdaderas es adoptar una actitud contraria. La distinción positivista entre hechos y valores, creencias y actitudes, sólo puede parecer plausible si suponemos que hay un vocabulario libre-de-valor que haga conmensurables a estos conjuntos de afirmaciones «factuales». Pero la ficción filosófica de que tenemos este vocabulario en la punta de la lengua es, desde un punto de vista educativo, desastrosa. Nos obliga a fingir que podemos dividirnos a nosotros mismos en conocedores de oraciones verdaderas por una parte y elegidores de vidas o acciones u obras de arte por la otra. Estas divisiones artificiales impiden que logremos enfocar la idea de edificación. O, más exactamente, nos llevan a pensar que la edificación no tiene nada que ver con las facultades racionales que se emplean en el discurso normal.

Así, pues, el esfuerzo de Gadamer por deshacerse de la imagen clásica del hombre-como-esencialmente-conocedor-de-esencias es, entre otras cosas, un esfuerzo por librarse de la distinción entre hecho y valor y, por tanto, por hacernos pensar que «descubrir los hechos» es sólo uno de los distintos proyectos de edificación. Ésta es la razón por la que Gadamer dedica tanto tiempo a acabar con las distinciones hechas por Kant entre cognición, moralidad y juicio estético [11]. No hay ninguna forma, por lo que yo sé, de ***argumentar*** en el tema de si se debe mantener en su sitio la «red» kantiana o más bien debe dejarse de lado. No hay ningún discurso filosófico normal que proporcione un terreno de conmensuración común a los que ven la ciencia y la edificación como, respectivamente, «racional»

[10] Véase la discusión de Heidegger sobre los «valores» en *Being and Time*, página 133, y lo que dice al respecto Sartre en *Being and Nothingness*, segunda parte, capítulo 1, sección 4. Compárense con los comentarios de Gadamer sobre Weber (*Truth and Method*, págs. 461 y ss.).

[11] Véase la polémica de Gadamer contra «la subjetivización de la estética» en la Tercera Crítica de Kant (*Truth and Method*, pág. 87) y compárese con los comentarios de Heidegger en «Letter on Humanism» hablando de las distinciones de Aristóteles entre física, lógica y ética (Heidegger, *Basic Writings*, ed. Krel [Nueva York, 1976], pág. 232).

e «irracional», y los que ven la búsqueda de la objetividad como una posibilidad entre otras que se deben tener en cuenta en la *wirkungsgeschichtliche Bewusstsein*. Si no existe ese terreno común, lo único que podemos hacer es indicar qué aspecto tiene el otro bando, desde nuestro propio punto de vista. Es decir, lo único que podemos hacer es ser hermenéuticos con la oposición —tratando de demostrar cómo las cosas extrañas o paradójicas u ofensivas que dicen van unidas al resto de lo que quieren decir, y qué impresión produce lo que dicen cuando se formula en nuestro propio idioma alternativo. Esta clase de hermenéutica de intención polémica aparece en los intentos de Heidegger y Derrida por demostrar la tradición.

2. Filosofía sistemática y filosofía edificante

El punto de vista hermenéutico, desde el cual la adquisición de la verdad ve reducida su importancia y se considera como un componente de la educación, sólo es posible cuando antes se ha adoptado otro punto de vista. La educación ha de comenzar con la aculturación. Por eso, para llegar a ser *gebildet* los primeros pasos han de ser la búsqueda de la objetividad y de la conciencia auto-consciente de las prácticas sociales en que consiste la objetividad. Debemos vernos primero como *en-soi* —descritos por aquellas afirmaciones que son objetivamente verdaderas en el juicio de nuestros semejantes— antes de que tenga sentido que nos veamos como *pour-soi*. De la misma manera, no podemos ser educados sin descubrir muchas cosas sobre las descripciones del mundo presentadas por nuestra cultura (por ejemplo, aprendiendo los resultados de las ciencias naturales). Quizá más adelante podamos dar menos valor a «estar en contacto con la realidad», pero sólo podremos permitírnoslo después de haber atravesado diversas etapas de conformidad, primero implícita y luego explícita y auto-consciente, con las normas de los discursos que se producen a nuestro alrededor.

Traigo a colación este punto trivial de que la educación —incluso la educación del revolucionario o del profeta— ha de comenzar con la conformidad con la aculturación y la conformidad simplemente para introducir un elemento de cautela en la afirmación «existencialista» de que la participación normal en el discurso normal es meramente un proyecto, una forma de estar en el mundo. La cautela consiste en decir que el discurso anormal y «existencial» depende siempre del discurso normal, que la posibilidad de la her-

menéutica depende siempre de la posibilidad (y quizá de la realidad) de la epistemología, y que la edificación utiliza siempre materiales proporcionados por la cultura de la época. Intentar el discurso anormal *de novo,* sin ser capaces de reconocer nuestra propia anormalidad, es una locura en el sentido más literal y terrible de la palabra. Insistir en ser hermenéutico donde podría valer la epistemología —volvernos incapaces de ver el discurso normal en términos de sus propios motivos, y capaces de verlo únicamente desde dentro de nuestro discurso anormal— no es una locura, pero refleja una deficiencia de educación. La adopción de la actitud «existencialista» hacia la objetividad y la racionalidad que es común a Sartre, Heidegger y Gadamer sólo tiene sentido si lo hacemos apartándonos conscientemente de una norma entendida correctamente. El «existencialismo» es un movimiento de pensamiento *intrínsecamente de reacción,* que sólo tiene sentido en oposición a la tradición. Quiero generalizar ahora este contraste entre filósofos cuya obra es esencialmente constructiva y aquellos cuya obra es esencialmente reactiva. Estableceré un contraste entre la filosofía que se centra en la epistemología y la filosofía que tiene su punto de partida en la desconfianza hacia las pretensiones de la epistemología. Se trata del contraste entre la filosofía «sistemática» y la «edificante».

En toda cultura suficientemente reflexiva hay quienes seleccionan un área, un conjunto de prácticas, y la consideran como el paradigma de la actividad humana. Luego tratan de demostrar cómo el resto de la cultura puede beneficiarse de este ejemplo. En el curso principal de la tradición filosófica occidental, este paradigma ha sido el *conocer* —poseer creencias verdaderas justificadas, o, todavía mejor, creencias tan intrínsecamente convincentes que hacen innecesaria la justificación. Dentro de esta corriente principal se han dado revoluciones filosóficas sucesivas promovidas por filósofos entusiasmados ante nuevas proezas cognitivas —por ejemplo, el redescubrimiento de Aristóteles, la mecánica de Galileo, el desarrollo de la historiografía auto-consciente en el siglo XIX, la biología darwiniana, la lógica matemática. La utilización de Aristóteles por Tomás de Aquino para conciliar a los Santos Padres, las críticas de Descartes y Hobbes a la escolástica, la idea de la Ilustración de que la lectura de Newton lleva naturalmente a la destitución de los tiranos, el evolucionismo de Spencer, el intento de Carnap de superar la metafísica mediante la lógica, son otros tantos intentos de remodelar el resto de la cultura inspirándose en el modelo del último logro cognitivo. Un filósofo característico de la «corriente principal» de Occidente diría: Ahora que tal-o-cual línea de investigación ha conseguido un

éxito tan sorprendente, vamos a rehacer toda la investigación y toda la cultura inspirándonos en su modelo, haciendo así posible que se impongan la objetividad y la racionalidad en áreas anteriormente oscurecidas por la convención, la superstición y la falta de una adecuada comprensión epistemológica de la capacidad del hombre para representar con exactitud a la naturaleza.

En la periferia de la historia de la filosofía moderna, aparecen figuras que, sin llegar a formar una «tradición», se parecen entre sí por su desconfianza ante la idea de que la esencia del hombre es ser un conocedor de esencias. Pertenecen a esta categoría Goethe, Kierkegaard, Santayana, William James, Dewey, el segundo Wittgesntein, y el segundo Heidegger. Muchas veces se les acusa de relativismo o de cinismo. Suelen formular dudas sobre el progreso, y especialmente sobre la última pretensión de que tal o cual disciplina ha conseguido por fin aclarar hasta tal punto la naturaleza del conocimiento humano que la razón se extenderá ahora por todos los confines de la actividad humana. Estos escritores han mantenido viva la idea de que, aun cuando tengamos una creencia verdadera justificada sobre todo lo que deseemos saber, quizá lo único que tengamos sea conformidad con las normas del momento. Han mantenido con vida la sensación historicista de que la «superstición» de este siglo ha sido el triunfo de la razón del pasado siglo, así como la sensación relativista de que el último vocabulario, tomado del descubrimiento científico más reciente, quizá no exprese representaciones privilegiadas de las esencias, sino que sea simplemente uno más del posible número infinito de vocabulario en que se puede describir el mundo.

Los filósofos que se mantienen dentro de la corriente principal son los filósofos a quienes llamaré «sistemáticos», y a los periféricos los llamaré «edificantes». Estos filósofos periféricos, pragmáticos, son escépticos en primer lugar *hacia la filosofía sistemática,* hacia todo el proyecto de conmensuración universal [12]. Los grandes pensadores edificantes, periféricos de nuestro tiempo son Dewey, Wittgenstein y Heidegger. En los tres casos resulta sumamente difícil interpretar su pensamiento como expresión de concepciones sobre los proble-

---

[12] Véase el pasaje de Anatole France, en «Garden of Epicurus», que cita Jacques Derrida al comienzo de su «La Mythologie Blanche» (en *Marges de la Philosophie* [París, 1972], pág. 250):

> ... los metafísicos, cuando inventan un nuevo lenguaje, son como los afiladores que aplican su piedra a las monedas y medallas en vez de a los cuchillos y tijeras. Borran el relieve, las inscripciones, los

mas filosóficos tradicionales, o como conjunto de propuestas constructivas para la filosofía en cuanto disciplina cooperativa y progresiva [13]. Se ríen de la imagen clásica del hombre, la imagen que contiene la filosofía sistemática, la búsqueda de la conmensuración universal en un vocabulario final. Insisten denodadamente en la afirmación holista de que las palabras toman sus significados de otras palabras más que en virtud de su carácter representativo, y en el corolario de que los vocabularios adquieren sus privilegios de los hombres que los usan más que de su transparencia a lo real [14].

La distinción entre filósofos sistemáticos y edificantes no es la misma que la distinción entre filósofos normales y filósofos revolucionarios. Esta última distinción pondría a Husserl, Russell, el segundo Wittgenstein y el segundo Heidegger en el mismo lado de la línea divisoria (el de los «revolucionarios»). Para mis intereses, lo que importa es distinguir entre dos clases de filósofos revolucionarios. Por una parte, hay filósofos revolucionarios —los que fundan nuevas escuelas dentro de las cuales se puede practicar la filosofía normal, profesionalizada— que consideran que la inconmensurabilidad de su nuevo vocabulario con el antiguo es un problema pasajero del que hay que echar la culpa a los fallos de sus predecesores y que se debe superar mediante la institucionalización de su propio vocabulario. Por otra parte, hay grandes filósofos que se asustan al pensar que su vocabulario pudiera llegar a institucionalizarse, o de que su obra pueda considerarse como conmensurable con la tradición.

---

retratos, y cuando ya no se puede ver en las monedas a Victoria, o a Guillermo o a la República Francesa, explican: estas monedas no tienen ahora nada de específicamente inglés o alemán o francés, pues las hemos alejado del tiempo y del espacio; ya no valen, por ejemplo, cinco francos, sino que tienen un valor incalculable y la zona en que pueden ser medio de cambio se ha ampliado hasta el infinito.

[13] Véase la comparación que hace Karl-Otto Apel del Wittgenstein y Heidegger en cuanto que han «cuestionado la metafísica occidental como disciplina teórica» *(Transformation der Philosophie* [Frankfurt, 1973], vol. 1, página 228). No he presentado interpretaciones de Dewey, Wittgenstein y Heidegger en apoyo de lo que vengo diciendo sobre ellos, pero he tratado de hacerlo en un trabajo sobre Wittgenstein titulado «Keeping Philosophy Pure» *(Yale Review* [Spring, 1976], págs. 336-356), en «Overcoming the Tradition: Heidegger and Dewey» *(Review of Metaphysics,* 30 [1976], págs. 280-305), y en «Dewey's Metaphysics», en *New Studies in the Philosophy of John Dewey,* edición de Steven M. Cahn (Hanover, N. H., 1977).

[14] Esta observación heideggeriana sobre el lenguaje aparece desarrollada ampliamente y con criterio en la obra de Derrida *La Voix et le Phénomène,* traducida al inglés por David Allison con el título *Speech and Phenomenon* (Evanston, 1973). Véase la comparación que hace Garver de Derrida y Wittgenstein en su «Introducción» a esta traducción.

Husserl y Russell (como Descartes y Kant) son del primer tipo. El segundo Wittgenstein y el segundo Heidegger (como Kierkegaard y Nietzsche) son del segundo [15]. Los grandes filósofos sistemáticos son constructivos y dan argumentos. Los grandes filósofos edificantes son reactivos y presentan sátiras, parodias, aforismos. Saben que su obra perderá vigencia cuando se pase el periodo contra el que estaban reaccionando. Son *intencionadamente* periféricos. Los grandes filósofos sistemáticos, como los grandes científicos, construyen para la eternidad. Los grandes filósofos edificantes destruyen en beneficio de su propia generación. Los filósofos sistemáticos quieren colocar su materia de estudio en el camino seguro de la ciencia. Los filósofos edificantes quieren dejar un espacio abierto a la sensación de admiración que a veces causan los poetas —admiración de que haya algo nuevo bajo el sol, algo que *no* sea una representación exacta de lo que ya estaba allí, algo que (al menos de momento) no se puede explicar y difícilmente se puede describir.

Sin embargo, la idea de un filósofo edificante constituye una paradoja. Platón definió al filósofo en oposición al poeta. El filósofo podía dar razones, argumentar sus opiniones, justificarse. Por eso, los filósofos sistemáticos defensores de la argumentación dicen de Nietzsche y Heidegger que, sean lo que sean, no son *filósofos*. Esta táctica de negar la condición de filósofos también la utilizan, lógicamente, los filósofos normales contra los filósofos revolucionarios. La utilizaron los pragmatistas contra los positivistas lógicos, los positivistas contra los «filósofos del lenguaje corriente», y será utilizada siempre que el cómodo profesionalismo esté en peligro. Pero en ese uso es simplemente un juego retórico que no nos dice más que se está proponiendo un discurso inconmensurable. Por el contrario, cuando se usa contra los filósofos edificantes la acusación tiene mordiente. El problema de un filósofo edificante es que en cuanto filósofo está en una profesión que consiste en presentar argumentos, mientras que lo que a él le gustaría es ofrecer otro conjunto de términos, *sin* decir que estos términos sean las representaciones exactas, recién encontradas, de las esencias (por ejemplo, de la esencia de la «filosofía» misma). Podría decirse que está violando no sólo las reglas de la filosofía normal (la filosofía de las escuelas de su época) sino una especie de meta-regla: la regla de que sólo se puede sugerir un

---

[15] La permanente fascinación del hombre que inventó toda la idea de la filosofía occidental —Platón— está en que todavía no sabemos qué clase de filósofo era. Aun cuando se deje de lado la *Séptima carta,* por apócrifa, el enigma se mantiene vivo por el hecho de que tras milenios de comentarios nadie sabe qué pasajes de los diálogos son bromas.

cambio de reglas porque se ha observado que las antiguas no encajan con la materia estudiada, que no son adecuadas a la realidad, que impiden la solución de los problemas eternos. Los filósofos edificantes, a diferencia de los filósofos sistemáticos revolucionarios, son los que son anormales en este meta-nivel. Se niegan a presentarse como si hubieran averiguado alguna verdad objetiva (por ejemplo, sobre qué es la filosofía). Se presentan como personas que hacen algo distinto de (y más importante que) ofrecer representaciones exactas de cómo son las cosas. Es más importante porque, dicen ellos, la misma idea de «representación exacta» no es la forma adecuada de pensar en lo que hace la filosofía. Pero luego continúan diciendo que esto no se debe a que «una búsqueda de representaciones exactas de... (por ejemplo, 'los rasgos más generales de la realidad' o 'la naturaleza del hombre')» sea una representación *in*exacta de la filosofía.

Mientras que los revolucionarios con menos pretensiones pueden permitirse tener opiniones sobre muchas cosas de las que las tenían también sus antecesores, los filósofos edificantes deben criticar la misma idea de tener una opinión, y al mismo tiempo evitar tener una opinión sobre tener opiniones [16]. Es una posición difícil, pero no imposible. Wittgenstein y Heidegger se defienden bastante bien. Una de las razones por las que consiguen arreglárselas tan bien es que no piensan que cuando decimos algo tengamos que estar expresando necesariamente una opinión sobre un tema. Podríamos estar *diciendo* algo simplemente —participando en una conversación más que colaborando en una investigación. Quizá el decir cosas no sea siempre decir cómo son las cosas. Quizá decir *eso* no sea un caso de decir cómo son las cosas. Los dos autores indican que vemos a las personas en cuanto que dicen cosas, cosas mejores o peores, sin verlas como si exteriorizaran representaciones internas de la realidad. Pero esto es sólo su cuña de entrada, pues entonces debemos dejar de vernos a nosotros mismos como que *vemos* esto, sin comenzar a vernos a nosotros mismos como si viéramos algo distinto. Debemos eliminar completamente de la conversación las metáforas visuales, y en especial las del espejo [17]. Para hacerlo hemos de entender

---

[16] La mejor exposición de esta dificultad, al menos de las que yo conozco, es la de Heidegger en «*Die Zeit des Weltbildes*» (traducido al inglés con el título «The Age of the World-View» por Marjorie Grene en *Boundary II* [1976]).

[17] Las obras recientes de Derrida son meditaciones sobre la forma de evitar estas metáforas. Igual que Heidegger en «Aus einem Gespräch von der Sprache zwischen einem Japaner und einem Fragenden» (en *Unterwegs zur Sprache* [Pfullingen, 1959]), Derrida juega en algunas ocasiones con la idea de la superioridad de las lenguas orientales y de la escritura ideográfica.

el habla no sólo como no exteriorización de las representaciones internas, sino como no representación en absoluto. Debemos renunciar a la idea de correspondencia de las oraciones y de pensamientos y ver las oraciones como si estuvieran conectadas con otras oraciones más que con el mundo. Debemos considerar el término «corresponde a cómo son las cosas» como un cumplido automático hecho al discurso normal que logra sus objetivos y no como una relación que se debe estudiar y a la que hay que aspirar durante todo el resto del discurso. Tratar de ampliar este cumplido a los hechos del discurso anormal es como felicitar a un juez por su sabia decisión dándole una buena propina: demuestra una falta de tacto. Pensar que Wittgenstein y Heidegger tienen opiniones sobre cómo son las cosas no es estar equivocados sobre cómo son las cosas, exactamente; es sólo mal gusto. Los coloca en una situación en la que no quieren estar, y en la que parecen ridículos.

Pero quizá *deban* parecer ridículos. ¿Cómo sabemos entonces cuándo adoptar una actitud discreta y cuándo insistir en la obligación moral de alguien a tener una opinión? Es como preguntar cómo sabemos cuándo la negativa de alguien a adoptar nuestras normas (por ejemplo, de organización social, prácticas sexuales o modales en la conversación) es un ultraje moral y cuándo es algo que debamos respetar (al menos provisionalmente). No sabemos estas cosas por referencia a principios generales. Por ejemplo, no sabemos por adelantado que si se pronuncia una determinada oración o se realiza un acto determinado, interrumpiremos la conversación o romperemos una relación personal, pues todo depende de qué es lo que lleva a ello. El ver a los filósofos edificantes como interlocutores en una conversación es una alternativa a verlos como si tuvieran opiniones sobre temas de interés común. Una forma de concebir la sabiduría como algo que no se ama igual que se ama un argumento, y cuya consecución no consiste en encontrar el vocabulario correcto para representar la esencia, es pensar en ella como la sabiduría práctica necesaria para participar en una conversación. Una forma de ver la filosofía edificante *como* amor a la sabiduría es verla como el intento de impedir que la conversación degenere en investigación, en un intercambio de opiniones. Los filósofos edificantes no pueden terminar la filosofía, pero pueden ayudar a impedir que llegue al sendero seguro de la ciencia.

3. Edificación, relativismo y verdad objetiva

Quiero ampliar ahora esta sugerencia de que la filosofía edificante aspira a mantener una conversación más que a descubrir la verdad, sirviéndome de ella para elaborar una réplica a la conocida acusación de «relativismo» dirigida a la subordinación de la verdad a la edificación. Voy a insistir en que la diferencia entre conversación e investigación es paralela a la distinción de Sartre entre pensar en sí mismo como *pour-soi* y como *en-soi* y, por tanto, que el papel cultural del filósofo edificante es ayudarnos a evitar el auto-engaño que se deriva de creer que nos conocemos a nosotros mismos conociendo un conjunto de hechos objetivos. En la sección siguiente, trataré de examinar la afirmación contraria. Lo que voy a decir es que el conductismo, naturalismo y fisicismo incondicionales que he estado recomendando en los capítulos precedentes nos ayudan a evitar el auto-engaño de pensar que poseemos una naturaleza profunda, oculta, metafísicamente significativa que nos hace «irreductiblemente» diferentes de los tinteros o de los átomos.

Los filósofos que tienen dudas sobre la epistemología tradicional son interpretados frecuentemente como si estuvieran cuestionando la idea de que sólo puede ser verdadera como máximo una de las teorías rivales incompatibles. Sin embargo, es difícil dar con alguien que ponga eso en duda. Cuando se dice, por ejemplo, que las «teorías de la verdad» coherentistas o pragmatistas admiten la posibilidad de que muchas teorías incompatibles satisfagan las condiciones establecidas para «la verdad», el coherentista o pragmatista suele replicar que esto demuestra solamente que no deberíamos tener motivos para elegir entre estos aspirantes a «la verdad». La conclusión que hay que sacar, dicen ellos, no es que hayan presentado análisis inadecuados de «verdadero», sino que hay algunos términos —por ejemplo, «la teoría verdadera», «lo que es correcto hacer»— que son, intuitiva y gramaticalmente, singulares, pero para los cuales no se puede establecer ningún conjunto de condiciones necesarias y suficientes que distingan a un referente único. Este hecho, dicen ellos, no tiene por qué causar sorpresa. Nadie piensa que haya condiciones necesarias y suficientes que distingan, por ejemplo, al referente único de «lo correcto en aquella situación embarazosa en que se encontró hubiera sido hacer esto», aunque se pueden presentar condiciones plausibles que abrevien una

lista de candidatos rivales incompatibles. ¿Por qué va a ser distinto con los referentes de «lo que deberíamos haber hecho en aquel terrible dilema moral» o «la Buena Vida para el hombre» o «de lo que está hecho realmente el mundo»?

Para ver al relativismo ocultándose en todo intento de formular condiciones de verdad o realidad o bondad que no trate de presentar condiciones singularmente individualizadoras, hemos de adoptar la noción «platónica» de los términos trascendentales de que hemos hablado más arriba (capítulo sexto, sección 6). Hemos de pensar que los verdaderos referentes de estos términos (la Verdad, lo Real, la Bondad) quizá no tengan ninguna conexión con las prácticas de justificación que prevalecen entre nosotros. El dilema creado por esta hipostatización platónica es que, por una parte, el filósofo ha de tratar de encontrar criterios para distinguir a estos referentes únicos, mientras que, por la otra, las únicas pistas sobre qué podrían ser estos criterios las proporciona la práctica actual (por ejemplo, la mejor corriente de pensamiento moral y científico del momento). Por eso, los filósofos se condenan a sí mismos a trabajar como Sísifo, pues en el momento en que se ha conseguido la explicación de un término trascendental recibe la etiqueta de «falacia naturalista», de confusión entre esencia y accidente [18]. Creo que tenemos una pista de la causa de esta obsesión contraproducente en el hecho de que incluso los filósofos que consideran la imposibilidad intuitiva de encontrar condiciones para «lo que es correcto hacer» como razón para rechazar los «valores objetivos» son reacios a considerar que la imposibilidad de encontrar condiciones individualizadoras para la única teoría verdadera del mundo sea razón para negar la «realidad física objetiva». Y deberían hacerlo, pues formalmente las dos nociones están en pie de igualdad. Las razones en favor y en contra de la adopción del enfoque de la «correspondencia» en la verdad moral son las mismas que las que se relacionan con la verdad sobre el mundo físico. La revelación se produce, pienso yo, cuando descubrimos que la excusa habitual del trato envidioso es que nos dejamos arrastrar por la realidad física y no por valores [19]. Pero, ¿qué tiene que ver el ser arrastrado con la objetividad, la representación exacta o la correspondencia? Nada,

---

[18] Sobre este punto, véase William Frankena y su obra clásica «The Naturalistic Fallacy», *Mind*, 68 (1939).

[19] Lo que parece ser una sensación de ser impulsado de un lado para otro por los valores, dicen ellos reductivamente, es simplemente la realidad física disfrazada (por ejemplo, distribuciones neurales o secreciones glandulares programadas por el condicionamiento paterno).

creo, a no ser que confundamos *contacto* con la realidad (relación causal, no intencional, no relativa a la descripción) con *hacer frente* a la realidad (describirla, explicarla, predecirla y modificarla —cosas todas ellas que hacemos bajo descripciones). El sentido en que la realidad física es «lo segundo» de Peirce —presión no mediada— no tiene nada que ver con el sentido en que sólo una de nuestras formas de describir, o de hacer frente a, la realidad física es «la única correcta». La falta de mediación se está confundiendo aquí con la exactitud de la mediación. La ausencia de descripción se confunde con un privilegio que va unido a cierta descripción. Sólo por obra de esta confusión se puede confundir la incapacidad de ofrecer condiciones individualizadoras para la única descripción verdadera de las cosas materiales con la insensibilidad a la inflexibilidad de las cosas.

Sartre nos ayuda a explicar por qué es tan frecuente esta confusión y por qué sus resultados están provistos de tanta seriedad moral. La idea de «la única forma de describir y explicar la realidad» que se supone contenida en nuestra «intuición» sobre el significado de «verdadero» es, para Sartre, precisamente la idea de tener una forma de describir y explicar que nos viene *impuesta* con la misma desconsideración con que chocan las piedras contra nuestros pies. O, para volver a metáforas visuales, es la idea de hacer que se nos desvele la realidad, no oscuramente en un cristal, sino como una especie inimaginable de inmediatez que haría superfluos el discurso y la descripción. Si pudiéramos convertir al conocimiento y hacerlo pasar de ser algo discursivo, algo conseguido por adaptaciones continuas de ideas o palabras, a ser algo tan inevitable como el ser arrastrados de una a otra parte, o el quedarse transpuestos ante una visión que nos deja sin habla, ya no tendríamos la responsabilidad de elegir entre palabras e ideas, teorías y vocabularios rivales. Este intento de sacudirse la responsabilidad es lo que Sartre describe como intento de convertirse en una cosa —en un *être-en-soi.* En las visiones del epistemólogo, esta noción incoherente toma la forma de visión de la consecución de la verdad como cuestión de *necesidad,* bien sea la necesidad «lógica» del trascendentalista o la necesidad «física» del epistemólogo «naturalizante» evolutivo. Desde el punto de vista de Sartre, el deseo de encontrar estas necesidades es el deseo de renunciar a la propia libertad para instaurar otra teoría o vocabulario alterativos. Por eso, el filósofo edificante que señala la incoherencia de este deseo es tachado de «relativista», de falta de seriedad moral, porque no comparte la común esperanza humana de que desaparezca la carga de la elección. Lo mismo

que del filósofo moral que entiende la virtud como auto-desarrollo aristotélico se piensa que no tiene interés por sus semejantes, el epistemólogo que es meramente conductista es tratado como quien no comparte la aspiración humana universal hacia la verdad objetiva.

Sartre contribuye a nuestra comprensión de las imágenes visuales que han planteado los problemas de la filosofía occidental ayudándonos a ver por qué estas imágenes siempre tratan de trascenderse a sí mismas. La idea de un Espejo de la Naturaleza totalmente limpio es la idea de un espejo que no podría distinguirse de lo que se ha reflejado y, por tanto, no sería un espejo en absoluto. La idea de un ser humano cuya mente es ese espejo limpio, y que *sabe* esto, es la imagen, como dice Sartre, de Dios. Este ser *no* se en frenta con algo extraño que le imponga la necesidad de elegir una actitud hacia él o una descripción del mismo. No tendría ninguna necesidad ni ninguna capacidad para elegir acciones o descripciones. Puede ser llamado «Dios» si pensamos en las ventajas de esta situación, o una «simple máquina» si pensamos en sus desventajas. Desde este punto de vista, el buscar la conmensuración en vez de limitarse a continuar la conversación —buscar una forma de hacer innecesaria una nueva redescripción encontrando una forma de reducir todas las *posibles* descripciones a una— es intentar escapar de la condición de hombre. Abandonar la idea de que la filosofía deba demostrar que todo discurso posible converge naturalmente en un consenso, igual que lo hace la investigación normal, sería abandonar la esperanza de ser algo más que meramente humano. Sería, por tanto, abandonar las ideas platónicas de Verdad y Realidad y Bondad en cuanto entidades que no pueden ser reflejadas ni siquiera confusamente por las actuales prácticas y creencias, y reincidir en el «relativismo» que supone que nuestras únicas ideas útiles de «verdadero» y «real» y «bueno» son extrapolaciones de esas prácticas y creencias.

Con esto vuelvo, finalmente, a la sugerencia con que he terminado la sección anterior —que el interés de la filosofía edificante es hacer que siga la conversación más que encontrar la verdad objetiva. Esta verdad, en la concepción que estoy defendiendo, es el resultado normal del discurso normal. La filosofía edificante no sólo es anormal sino reactiva, ya que sólo tiene sentido como protesta contra los intentos de cortar la conversación haciendo propuestas de conmensuración universal mediante la hipostatización de un conjunto privilegiado de descripciones. El peligro que trata de evitar el discurso edificante es que algún vocabulario dado, alguna forma en que las personas pudieran llegar a pensar en sí mismas, les decida

a pensar que de ahora en adelante todo discurso podría ser, o debería ser, discurso normal. La paralización resultante de la cultura sería, a los ojos de los filósofos edificantes, la deshumanización de los seres humanos. Los filósofos edificantes están, por consiguiente, de acuerdo con la opción de Lessing por la aspiración infinita a conseguir la verdad y no por «toda la Verdad»[20]. Para el filósofo edificante la misma idea de encontrarse con «toda la Verdad» resulta en sí absurda, pues también es absurda la idea Platónica de la Verdad misma. Es absurda tanto como noción de verdad sobre la realidad que no es sobre la realidad-bajo-una-determinada-descripción, como en forma de noción de verdad sobre la realidad bajo una descripción privilegiada que haga innecesarias todas las demás descripciones porque es conmensurable con cada una de ellas.

Pensar que mantener una conversación es una meta suficiente de la filosofía, ver la sabiduría como si consistiera en la capacidad de mantener una conversación, es considerar a los seres humanos como generadores de nuevas descripciones más que como seres de quienes se espera que sean capaces de describir con exactitud. Pensar que la meta de la filosofía es la verdad —a saber, la verdad sobre los términos que constituyen una conmensuración última para todas las investigaciones y actividades humanas— es ver a los seres humanos como objetos más que como sujetos, como existentes *en-soi* más que como *pour-soi* y *en-soi,* como objetos descritos y sujetos que describen simultáneamente. Pensar que la filosofía nos permitirá ver al sujeto que describe como si fuera él mismo un tipo de objeto descrito es pensar que todas las posibles descripciones pueden volverse conmensurables con ayuda de un solo vocabulario descriptivo —el de la filosofía misma—. Sólo si tuviéramos esta idea de una descripción universal podríamos identificar a los seres-humanos-bajo-una-determinada-descripción con la «esencia» del hombre. Sólo con esa idea tendría sentido la idea de un hombre *que tiene* una esencia, sea concebida o no esa esencia como el conocimiento de las esencias. Así, pues, ni siquiera diciendo que el hombre es a la vez sujeto y objeto, *pour-soi* y *en-soi,* estamos captando nuestra esencia. No escapamos del Platonismo diciendo que «nuestra esencia es no tener ninguna esencia» si luego tratamos de servirnos de esta idea como base para un intento constructivo y sistemático de averiguar nuevas verdades sobre los seres humanos.

Esa es la razón por la que el «existencialismo» —y, más en ge-

[20] Kierkegaard hizo de esta opción el prototipo de su propia opción por la «subjetividad» más que por el «sistema». Cfr. *Concluding Unscientific Postscript,* traducción de David Swenson y Walter Lowrie (Princeton, 1941), pág. 97.

neral, la filosofía edificante— *sólo* puede ser reactivo, la razón por la que se engaña a sí mismo siempre que intenta hacer algo más que hacer que la conversación siga por nuevos caminos. Estos nuevos caminos pueden engendrar, quizá, nuevos discursos normales, nuevas ciencias, nuevos programas de investigación filosófica y, por tanto, nuevas verdades objetivas. Pero no constituye el sentido de la filosofía edificante, siendo sólo subproductos accidentales. Su sentido es siempre el mismo —realizar la función social que Dewey llamaba «romper la costra de la convención», impidiendo que el hombre se engañe a sí mismo con la idea de que se conoce a sí mismo o que conoce alguna otra cosa, como no sea bajo descripciones opcionales.

4. Edificación y naturalismo

En el capítulo séptimo he insistido en que sería buena idea prescindir de la distinción espíritu-naturaleza, entendida como división entre seres humanos y otras cosas, o entre dos partes de los seres humanos, que correspondería a la distinción entre hermenéutica y epistemología. Quiero volver a examinar este tema, para subrayar que las doctrinas «existencialistas» de que vengo hablando son compatibles con el conductismo y materialismo que he defendido en capítulos anteriores. Los filósofos a quienes les gustaría ser simultáneamente sistemáticos y edificantes han estimado que son incompatibles y, por tanto, han propuesto la forma en que nuestro sentido de nosotros mismos como *pour-soi,* como capaz de reflexión, como sujetos que eligen vocabularios alternativos, podría convertirse en uno de los objetos estudiados por la filosofía.

Gran parte de la filosofía reciente —bajo los auspicios de la «fenomenología» o de la «hermenéutica», o de ambas— ha coqueteado con esta idea tan desafortunada. Por ejemplo, Habermas y Apel han insinuado formas de crear un nuevo tipo de punto de vista trascendental, que nos permitiría hacer algo parecido a lo que intentó hacer Kant, pero sin incurrir ni en el cientismo ni en el historicismo. Asimismo, la mayoría de los filósofos que consideran a Marx, a Freud o a ambos como figuras que deben introducirse en el «cauce principal» de la filosofía han tratado de desarrollar sistemas cuasi-epistemológicos que se centran en torno al fenómeno que Marx y Freud ponen de relieve —el cambio en la conducta que se produce como consecuencia del cambio en la autodescripción. Estos filósofos piensan que la epistemología tradicional

se ha dedicado a «objetivizar» a los seres humanos, y esperan que llegue un sucesor de la epistemología que haga por la «reflexión» lo que la tradición ha hecho por el «conocimiento objetivizante».

He insistido varias veces en que no deberíamos intentar tener un sucesor de la epistemología, sino más bien tratar de liberarnos de la idea de que la filosofía tiene que estar centrada en torno al descubrimiento de un marco permanente de investigación. Desde mi punto de vista, el intento de desarrollar una «pragmática universal» o una «hermenéutica trascendental» resulta muy sospechoso. Parece que promete justamente lo que Sartre nos dice que no vamos a tener —una forma de ver la libertad como naturaleza (o, con un lenguaje menos críptico, una forma de ver nuestra creación de vocabularios y la elección entre los mismos de la misma forma «normal» que nos vemos a nosotros mismos *dentro* de uno de esos vocabularios). Estos intentos comienzan considerando la búsqueda de conocimiento objetivo a través del discurso normal tal como he indicado que debería considerarse —como un elemento dentro de la edificación. Pero luego pasan a formular afirmaciones más ambiciosas. Puede servir de ejemplo el siguiente pasaje de Habermas:

> ... las funciones que tiene el conocimiento en los contextos universales de la vida práctica sólo se pueden analizar con éxito en el marco de una filosofía trascendental reformulada. Esto, dicho sea de paso, no implica una crítica empirista a la afirmación de la verdad absoluta. Mientras los intereses cognitivos puedan ser identificados y analizados mediante la reflexión sobre la lógica de la investigación en las ciencias naturales y culturales, pueden reclamar legítimamente un *status* «trascendental». Asumen un *status* «empírico» en el momento en que son analizados como resultado de la historia natural —analizados, por así decirlo, en términos de la antropología cultural [21].

Lo que yo quiero decir, por el contrario, es que no tiene sentido tratar de encontrar una forma sinóptica general de «analizar» las «funciones que tiene el conocimiento en los contextos universales de la vida práctica», y que lo único que necesitamos es la antro-

[21] Jürgen Habermas, «Nachwort» a la segunda edición de *Erkenntnis und Interesse* (Frankfurt: Surkamp, 1973), pág. 410; traducido al inglés con el título «A Postscript to *Knowledge and Human Interests*», por Christian Lenhardt, en *Philosophy of the Social Sciences*, 3 (1973), pág. 181. Puede verse una crítica a la línea seguida por Habermas —crítica que es paralela a la mía— en Michael Theunissen, *Gesellschaft und Geschichte: Zur Kritischen Theorie* (Berlín, 1969), págs. 20 y ss. (Debo la referencia a Theunissen y a Raymond Geuss.)

pología cultural (en un sentido amplio que incluye la historia intelectual).

Habermas y otros autores a quienes mueven los mismos motivos consideran que la sugerencia de que basta con la investigación empírica lleva consigo una «ilusión objetivista». Suelen ver el pragmatismo deweyano, y el «realismo científico» de Sellars y Feyerabend, como productos de una epistemología inadecuada. En mi opinión, la gran virtud de Dewey, Sellars y Feyerabend es que señalan el camino hacia (y en parte ejemplifican) una clase no epistemológica de filosofía y, que por tanto, renuncian a toda esperanza de lo «trascendental». Habermas dice que el que una teoría «tenga bases trascendentales» significa que

> adquiera familiaridad con el ámbito de condiciones subjetivas inevitables que hacen posible la teoría *y* al mismo tiempo la limitan, pues esta forma de corroboración trascendental tiende siempre a criticar una autocomprensión de sí mismo excesivamente segura de sí misma [22].

Más en concreto, este exceso de confianza consiste en pensar que

> puede haber eso que se llama veracidad con la realidad en el sentido postulado por el realismo filosófico. Las teorías de la correspondencia de la verdad suelen hipostatizar los hechos en forma de realidad que están en el mundo. La intención y lógica interna de una epistemología que reflexiona sobre las condiciones de la experiencia posible en cuanto tal es descubrir las ilusiones objetivistas de semejante concepción. Todas y cada una de las formas de filosofía trascendental pretenden identificar las condiciones de la objetividad de la experiencia analizando la estructura categórica de los objetos de la experiencia posible [23].

Pero Dewey, Wittgenstein, Sellars, Kuhn y los otros protagonistas del presente libro tienen todos sus formas propias de desacreditar «la veracidad con la realidad en el sentido postulado por el realismo filosófico» y ninguno de ellos piensa que esto deba hacerse «analizando la estructura categórica de los objetos de la experiencia posible».

La idea de que sólo podemos soslayar las reducciones positivistas y el realismo filosófico demasiado confiado adoptando algo parecido al punto de vista transcendental de Kant me parece el error básico de programas como el de Habermas (así como de la idea de Husserl de una «fenomenología del mundo de la vida» que describa

[22] Habermas, «Nachwort», pág. 411; traducción inglesa, pág. 182.
[23] *Ibíd.*, págs. 408-409; traducción inglesa, pág. 180.

a las personas en forma «previa» a la ofrecida por la ciencia). Lo que hace falta para conseguir estos propósitos laudables no es la distinción «epistemológica» de Kant entre los puntos de vista transcendental y empírico, sino más bien su distinción «existencialista» entre las personas como egos empíricos y como agentes morales [34]. El discurso científico normal puede verse siempre de formas diferentes —como la búsqueda fructífera de la verdad objetiva, o como un discurso más entre otros, uno más entre los muchos proyectos en que nos comprometemos. El primer punto de vista está de acuerdo con la práctica normal de la ciencia normal. En ella no se presentan cuestiones de opción moral o de edificación, pues ya desde antes se ha impuesto el compromiso tácito y «confiado» de buscar la verdad objetiva sobre el tema en cuestión. Desde el último punto de vista nos formulamos preguntas como «¿Qué sentido tiene?», «¿Qué conclusión hay que sacar de nuestro conocimiento de cómo actuamos nosotros, y el resto de la naturaleza?» o «¿Qué vamos a hacer con nosotros mismos ahora que conocemos las leyes de nuestra propia conducta?»

El error principal de la filosofía sistemática ha sido siempre la idea de que estas preguntas deben contestarse mediante un discurso nuevo («metafísico» o «transcendental») descriptivo o explicativo (que trate, por ejemplo, de «hombre», «espíritu», o «lenguaje»). Este intento de contestar a las cuestiones de justificación descubriendo nuevas verdades objetivas, de responder a la búsqueda de justificaciones por parte del agente moral con descripciones de un dominio privilegiado, es la forma especial de mala fe del filósofo —su forma especial de sustituir la elección moral por la seudo-cognición. La grandeza de Kant estuvo en haber logrado ver a través de la forma «metafísica» de este intento, y haber destruido la concepción tradicional de la razón para dejar lugar a la fe moral. Kant nos dio una forma de ver la verdad científica como algo que nunca podría ofrecernos una respuesta a nuestra búsqueda de sentido, una justificación, una forma de afirmar que nuestra decisión moral sobre lo que habría que hacer está basada en el *conocimiento* de la natu-

[34] Wilfrid Sellars utiliza positivamente esta última distinción kantiana cuando insiste en que la personalidad es cuestión de «ser uno de nosotros», de caer dentro del ámbito de los imperativos prácticos de la forma «Si eso fuera todo...», más que un rasgo de ciertos organismos que se podría aislar con medios empíricos. He manifestado esta pretensión varias veces a lo largo del libro, especialmente en el capítulo cuarto, sección 4. Sobre el uso de la misma en Sellars véase *Science and Metaphysics* (Londres y Nueva York, 1968), capítulo 7, y el ensayo «Science and Ethics», en su *Philosophical Perspectives* (Springfield, Ill., 1967).

raleza del mundo. Por desgracia, Kant formuló su diganóstico de la ciencia en términos del descubrimiento de «condiciones subjetivas inevitables», que serían reveladas mediante la reflexión sobre la investigación científica. Fue igualmente desafortunado el pensar que había realmente un procedimiento de decisión para los dilemas morales (aunque no basado en el *conocimiento,* pues nuestra captación del imperativo categórico no es una *cognición)* [25]. Por eso, creó nuevas formas de mala fe filosófica —sustituyendo los intentos «metafísicos» de encontrar un mundo en alguna otra parte por los intentos «transcendentales» de encontrar el verdadero yo de cada uno. Al identificar tácitamente al agente moral con el ego transcendental constituyente, abrió la puerta a los intentos post-kantianos cada vez más complicados de reducir la libertad a la naturaleza, la elección al conocimiento, el *pour-soi* a *en-soi.* Éste es el camino que estoy intentando cerrar rehaciendo las distinciones ahistóricas y permanentes entre naturaleza y espíritu, «ciencia objetivizante» y reflexión, epistemología y hermenéutica, en términos de distinciones históricas y temporales entre lo que resulta familiar y lo que no, entre lo normal y lo anormal. En cuanto a esta forma de tratar a las mencionadas distinciones conviene que las veamos no como si dividieran dos áreas de investigación, sino como la distinción entre investigación y algo que *no* es investigación, sino más bien como el interrogatorio inicial a partir del cual pueden sugir (o no) investigaciones —nuevos discursos normales.

Por decirlo de otra manera que puede servir para hacer resaltar sus conexiones con el naturalismo, lo que estoy señalando es que los positivistas tenían toda la razón al considerar imperioso extirpar la metafísica, cuando «metafísica» significa el intento de ofrecer el conocimiento de lo que la ciencia no puede conocer. Esto es sólo un intento de encontrar un discurso que combine las ventajas de la normalidad con las de la anormalidad —la seguridad intersubjetiva de la verdad objetiva combinada con el carácter edificante de una exigencia moral injustificable, pero incondicional. El deseo de colocar a la filosofía en la senda segura de la ciencia es el deseo de combinar el proyecto platónico de elección moral en cuanto determinación de verdades objetivas sobre un objeto de una clase

[25] Véase la distinción de Kant entre conocimiento y creencia necesaria en *K. d. r. V.,* A824-B852 y ss., y especialmente su utilización de *Unternehmung* como sinónimo de la segunda. Esta sección de la Primera Crítica me parece que es la que más sentido da al famoso pasaje sobre la negación de la razón para dejar lugar a la fe, en Bxxx. En muchos otros puntos, sin embargo, Kant habla incoherentemente de la razón práctica como si proporcionara una ampliación de nuestro *conocimiento.*

especial (la Idea del Bien) con el acuerdo intersubjetivo y democrático sobre los objetos que se da en la ciencia normal[26]. La filosofía que no tuviera nada de edificante, que fuera totalmente irrelevante para opciones morales como creer o no creer en Dios no se tendría por *filosofía,* sino sólo como una clase especial de ciencia. Por eso, en el mismo momento en que un programa de colocar a la filosofía en el sendero seguro de la ciencia consigue su objetivo, lo único que hace es convertir a la filosofía en una aburrida especialidad académica. La filosofía sistemática existe gracias a un perpetuo nadar entre dos aguas, a tener una pierna en cada lado del abismo que separa la descripción y la justificación, la cognición y la elección, el captar los hechos como son y decirnos cómo vivir.

Una vez visto esto, podemos apreciar con más claridad por qué la epistemología apareció como la esencia de la filosofía sistemática. La epistemología es el intento de ver los patrones de justificación dentro del discurso normal como algo *más* que esos patrones. Es el intento de verlos conectados a algo que exige un compromiso moral —la Realidad, la Verdad, la Objetividad, la Razón. Por el contrario, ser conductista en epistemología es mirar bifocalmente el discurso científico moral de nuestros días, como patrones adoptados por varias razones históricas y como la consecución de la verdad objetiva, donde «verdad objetiva» es ni más ni menos que la mejor idea que tenemos en la actualidad sobre cómo explicar lo que está ocurriendo. Desde el punto de vista del conductismo epistemológico, la única verdad en la afirmación de Habermas de que la investigación científica es posibilitada, y limitada, por «condiciones subjetivas inevitables» es que dicha investigación se hace posible por la adopción de prácticas de justificación y que tales prácticas tienen alternativas posibles. Pero estas «condiciones subjetivas» no son en ningún sentido las condiciones «inevitables» que se pueden descubrir mediante la «reflexión sobre la lógica de la investigación». Son simplemente los hechos sobre los que una determinada sociedad, o profesión, u otro grupo, considera que es buena base para afirmaciones de cierto tipo. Estas matrices disciplinares se estudian con los métodos empírico-más-hermenéutico de la «antropología cultural». Desde el punto de vista del grupo en cuestión estas condiciones subjetivas son una combinación de imperativos prácticos de sentido común (por ejem-

[26] Los propios positivistas sucumbieron rápidamente a este impulso. Aun insistiendo en que las cuestiones morales eran no-cognitivas, pensaron en dar *status* cuasi-científico a sus ataques moralistas a la filosofía tradicional —convirtiéndose así en motivos de críticas auto-referenciales en relación con su uso «emotivo» de lo «no-cognitivo».

plo, tabúes tribales, Métodos de Mill) con la teoría actualmente en vigor sobre el tema. Desde el punto de vista del historiador de las ideas o del antropólogo son los hechos empíricos sobre las creencias, deseos y prácticas de un cierto grupo de seres humanos. Son puntos de vista incompatibles, en el sentido de que no podemos adoptarlos simultáneamente. Pero no hay ninguna razón ni ninguna necesidad de subsumir los dos en una síntesis superior. El grupo en cuestión puede pasar de un punto de vista al otro («objetivizando» así sus egos pasados mediante un proceso de «reflexión» y haciendo oraciones que sean verdaderas de sus egos presentes). Pero esto no es un proceso misterioso que exija una nueva comprensión del conocimiento humano. Es el hecho de dominio público de que las personas pueden llegar a tener dudas sobre lo que están haciendo y, por tanto, comenzar a disertar en formas inconmensurables con las que utilizaban anteriormente.

Esto sirve también para los discursos nuevos más espectaculares y perturbadores. Cuando filósofos edificantes como Marx, Freud y Sartre ofrecen nuevas explicaciones de nuestros patrones habituales de justificación de nuestras acciones y afirmaciones, y cuando estas explicaciones son incorporadas e integradas en nuestras vidas, tenemos ejemplos sorprendentes del fenómeno de que la reflexión cambia el vocabulario y la conducta. Pero como he argumentado en el capítulo séptimo, este fenómeno no requiere ninguna nueva comprensión de la construcción de teorías o de la confirmación de teorías. Decir que nos hemos cambiado a nosotros mismos interiorizando una auto-descripción nueva (utilizando términos como «intelectual burgués» o «auto-destructivo» o «que se auto-engaña») tiene mucho de verdad. Pero esto no tiene más de sorprendente que el hecho de que los hombres cambiaran los datos de la botánica por hibridación, que a su vez fue hecho posible por la teoría botánica, o que cambiaran sus propias vidas inventando bombas y vacunas. La meditación sobre la posibilidad de tales cambios, como la lectura de obras de ciencia ficción, nos ayuda a superar la confianza en sí mismo del «realismo filosófico». Pero esta meditación no necesita verse complementada por una explicación transcendental de la naturaleza de la reflexión. Lo único que es necesario es la innovación edificante del hecho o posibilidad de discursos anormales, que minarían nuestra confianza en el conocimiento obtenido mediante discursos normales. La reprochable confianza en sí mismo a que nos estamos refiriendo es simplemente la tendencia del discurso normal a bloquear el flujo de la conversación, presentándose a sí mismo como si ofreciera el vocabulario canónico para la discusión de un tema

determinado —y, más en especial, la tendencia de la filosofía epistemológicamente centrada a bloquear el camino presentándose como el vocabulario final de conmensuración para todo discurso racional *posible*. La confianza en sí mismo perteneciente al primer tipo, es decir, la confianza limitada, la echan abajo los filósofos edificantes que ponen en duda la misma idea de conmensuración universal y de filosofía sistemática.

Aun a riesgo de repetirme más de lo necesario, quiero insistir una vez más en que la distinción entre discurso normal y anormal no coincide con ninguna distinción del objeto material (por ejemplo, naturaleza frente a historia, o hechos frente a valores), método (por ejemplo, objetivación frente a reflexión), facultad (por ejemplo, razón frente a imaginación) o cualquiera otra de las distinciones que ha utilizado la filosofía sistemática para hacer que el sentido del mundo consista en la verdad objetiva sobre alguna parte o rasgo del mundo que antes habían pasado desapercibidos. El discurso anormal puede versar sobre *cualquier cosa*, lo mismo que no hay nada que no pueda resultar edificante o que no se pueda sistematizar. He examinado la relación entre ciencia natural y otras disciplinas simplemente porque, desde la época de Descartes y Hobbes, el motivo clásico que ha impulsado a filosofar ha sido la suposición de que el discurso científico era discurso normal y que todo otro discurso debía estar inspirado en ese modelo. Sin embargo, en el momento en que dejamos de lado esta suposición, podemos también dejar de lado los diversos anti-naturalismos de que me vengo lamentando. Más concretamente, podemos afirmar todo lo que sigue:

Toda habla, pensamiento, teoría, poema, composición y filosofía resultará completamente previsible en términos puramente naturalistas. Alguna explicación tipo átomos-y-vacío aplicada a los microprocesos que ocurren dentro de los seres humanos individuales permitirá la predicción de todo sonido o inscripción que se llegue a producir. No hay espíritus.

Nadie será capaz de predecir sus propias acciones, pensamientos, teorías, poemas, etc., antes de tomar una decisión sobre ellos o inventarlos. (Esto no constituye una observación interesante sobre la extraña naturaleza de los seres humanos, sino más bien una consecuencia sin importancia de lo que significa «decidir» o «inventar».) Por eso, no existe ninguna esperanza (ni peligro) de que la cognición de uno mismo como *en soi* le obligue a dejar de existir *pour-soi*.

El conjunto completo de leyes que permitan hacer estas predicciones, más las descripciones completas (en términos de átomos-y-vacío) de todos los seres humanos, no llegaría a constituir toda la

«verdad objetiva» sobre los seres humanos, ni todo el conjunto de predicciones verdaderas sobre ellos. Seguiría habiendo tanto otros conjuntos distintos de tales verdades objetivas (algunas útiles en orden a la predicción, otras no) como vocabularios inconmensurables, en los que pudiera realizarse la investigación normal sobre los seres humanos (por ejemplo, todos aquellos vocabularios dentro de los cuales atribuimos creencias y deseos, virtudes y belleza).

La inconmensurabilidad implica irreductibilidad, pero no incompatibilidad, por lo que el que no se puedan «reducir» estos diversos vocabularios al de la ciencia de átomos-y-vacío, el del último nivel, no arroja ninguna duda sobre su *status* cognitivo, sobre el *status* metafísico de sus objetos. (Esto se aplica tanto al valor estético de los poemas como a las creencias de las personas, tanto a las virtudes como a las voliciones.)

La reunión, *per impossibile,* de todas estas verdades objetivas no sería necesariamente edificante. Podría ser la imagen de un mundo sin sentido, sin moral. El que a un individuo le parezca que apunta a una moral, es algo que depende de ese individuo. Sería verdadero o falso que a él se lo pareciera, o no se lo pareciera. Pero no sería objetivamente verdadero o falso que «tuviera realmente», o que no tuviera, un sentido o una moral. Se puede predecir si su conocimiento del mundo le deja un sentido de lo que debe hacer con o en el mundo, pero no es previsible si *debería* dejárselo o no.

El miedo a la ciencia, al «cientismo», al «naturalismo», a la auto-objetivación, a que el exceso de conocimiento llegue a convertirnos en una cosa más que en una persona, es el temor a que todo discurso se convierta en discurso normal. Es decir, es el temor a que haya respuestas objetivamente verdaderas o falsas a todas las preguntas que podamos formular, de manera que la valía del hombre consistiera en conocer verdades, y la virtud humana fuera meramente la creencia verdadera justificada. Esto resulta aterrador, pues excluye la posibilidad de que haya algo nuevo bajo el sol, de que la vida humana sea poética y no meramente contemplativa.

Pero los peligros para el discurso anormal no vienen de la ciencia o de la filosofía naturalista. Proceden de la escasez de alimentos y de la policía secreta. Si hubiera tiempo libre y bibliotecas, la conversación que inició Platón no terminaría en la auto-objetivación —no porque algunos aspectos del mundo, o de los seres humanos, se libren de ser objetos de la investigación científica, sino simplemente porque la conversación libre y relajada produce discurso anormal en el momento en que hace saltar chispas.

5. La filosofía en la conversación de la Humanidad

Terminaré el libro con una alusión al famoso título de Oakeshott [27], pues capta el tono en que, pienso yo, debe estudiarse la filosofía. Gran parte de lo que he dicho sobre la epistemología y sus posibles sucesores es un intento de extraer algunos corolarios de la doctrina de Sellars, según el cual

> al caracterizar un episodio o estado como de *conocer,* no estamos ofreciendo una descripción empírica de tal episodio o estado; lo estamos colocando en el espacio lógico de las razones, de justificar y ser capaz de justificar lo que se dice [28].

Si consideramos que el conocer no es algo que tenga una esencia, que debe ser descrita por científicos o filósofos, sino más bien como un derecho, según las normas en vigor, a creer, estamos avanzando hacia una comprensión de la *conversación* como el contexto último dentro del cual se debe entender el conocimiento. Nuestro centro de atención pasa de la relación entre seres humanos y objetos de su investigación a la relación entre criterios alternativos de justificación, y de ahí a los cambios reales en dichos criterios que integran la historia intelectual. Esto nos lleva a apreciar la descripción que hace Sellars de su héroe mítico Jones, el hombre que inventó el Espejo de la Naturaleza y con ello hizo posible la filosofía moderna:

> ¿No reconoce el lector a Jones como el Hombre mismo en la mitad de su viaje desde los gruñidos y gemidos de la caverna al discurso sutil y polidimensional de la sala de estar, del laboratorio y del estudio, al lenguaje de Henry y William James, de Einstein y de los filósofos que, en sus esfuerzos por salir del discurso para llegar a un αρχή más allá del discurso, han aportado la dimensión más curiosa de todas? (pág. 196).

En este libro he presentado una especie de prolegómeno para una historia de la filosofía centrada en la epistemología como si fuera un episodio dentro de la historia de la cultura europea. Esta filosofía se remonta a los griegos, y penetra de soslayo en todo tipo de disciplinas no filosóficas que, en uno u otro momento, se han propuesto a sí mismas como sustitutos de la epistemología y, por tanto,

[27] Cfr. Michael Oakeshott, «The Voice of Poetry in the Conversation of Mankind», en su *Rationalism and Politics* (Nueva York, 1975).

[28] Wilfrid Sellars, *Science, Perception and Reality* (Londres y Nueva York, 1963), pág. 169.

de la filosofía. Por eso, el episodio en cuestión no debe identificarse sin más con la «filosofía moderna», en el sentido de la secuencia de los manuales clásicos formada por grandes filósofos que van desde Descartes a Russell y Husserl. Pero es esa secuencia, no obstante, el lugar donde más explícita resulta la búsqueda de fundamentos para el conocimiento. Por eso, la mayoría de los esfuerzos que he realizado por desmontar la imagen del Espejo de la Naturaleza han tenido que ver con estos filósofos. He tratado de demostrar cómo su deseo de escapar a un αρχή más allá del discurso está arraigado en el deseo de ver las prácticas sociales de justificación como algo más que si fueran sólo tales prácticas. No obstante, me he concentrado principalmente en las expresiones de este deseo en las publicaciones más recientes de la filosofía analítica. El resultado no es, por consiguiente, más que un prolegómeno. Un tratamiento histórico adecuado requeriría conocimientos y habilidades que yo no poseo. Pero tengo la esperanza de que el prolegómeno haya servido para hacer ver que los problemas de la filosofía contemporánea constituyen hechos ocurridos en una fase determinada de una conversación —conversación que en algunos momentos no sabía nada de estas cuestiones y puede que vuelva a no saber nada de las mismas.

El hecho de que podamos continuar la conversación que inició Platón sin tratar los temas que Platón quería tratar, sirve para ilustrar la diferencia entre tratar a la filosofía como una voz dentro de un conversación y tratarla como una asignatura, una *Fach,* un campo de investigación profesional. La conversación iniciada por Platón se ha ampliado con más voces de las que Platón se habría atrevido a imaginar y, por tanto, con temas de los que él no sabía nada. Una «materia» —la astrología, la física, la filosofía clásica, el diseño de muebles— puede experimentar revoluciones, pero obtiene su autoimagen de su estado actual, y su historia tiene que escribirse necesariamente en forma de narración de su maduración gradual. Ésta es la forma más frecuente de escribir la historia de la filosofía, y yo no puedo pretender haberlo evitado del todo al esbozar la clase de historia que se debe escribir. Pero espero haber demostrado cómo es posible considerar las cuestiones de que se ocupan actualmente los filósofos, y de las que ellos consideran que se ha ocupado siempre (quizá sin saberlo) la filosofía, como resultados de accidentes históricos, como giros que ha tomado la conversación [29]. Durante

[29] Dos escritores recientes —Michael Foucault y Harold Bloom— dan este sentido a la factualidad bruta de los orígenes históricos fundamentales en su obra. Cfr. Bloom, *A Map of Misreading* (Nueva York, 1975), pág. 33:

mucho tiempo ha seguido este derrotero, pero quizá adopte otra dirección sin que por ello los seres humanos pierdan su razón, ni pierdan contacto con «los problemas reales».

El interés conversacional de la filosofía en cuanto materia, o de algún filósofo individual dotado de talento, ha variado y seguirá variando en formas imprevisibles que dependen de contingencias. Estas contingencias van desde lo que ocurre en la física a lo que ocurre en la política. Las líneas divisorias entre disciplinas se difuminarán y cambiarán, y aparecerán nuevas disciplinas, en las formas ilustradas por el intento logrado de Galileo de crear, en el siglo XVII, «cuestiones puramente científicas». Las nociones de «significación filosófica» y de «cuestión puramente filosófica», tal como se usan en la actualidad, sólo adquirieron sentido aproximadamente en la época de Kant. Nuestra sensación post-kantiana de que la epistemología o algún sucesor suyo está en el centro de la filosofía (y que la filosofía moral, la estética y la filosofía, por ejemplo, son en cierto sentido secundarias) es reflejo del hecho de que la auto-imagen del filósofo profesional depende de su preocupación profesional por la imagen del Espejo de la Naturaleza. Sin la suposición kantiana de que el filósofo puede decidir *quaestiones juris* relacionadas con las pretensiones del resto de la cultura, esta auto-imagen se viene abajo. Esa suposición depende de la idea de que existe eso que se llama comprensión de la esencia del conocimiento —hacer lo que Sellars nos dice que no podemos.

Abandonar la idea del filósofo en cuanto persona que conoce algo sobre el conocer que ningún otro conoce tan bien sería abandonar la idea de que su voz tiene derecho preferente a la atención de los demás participantes en la conversación. Sería también aban-

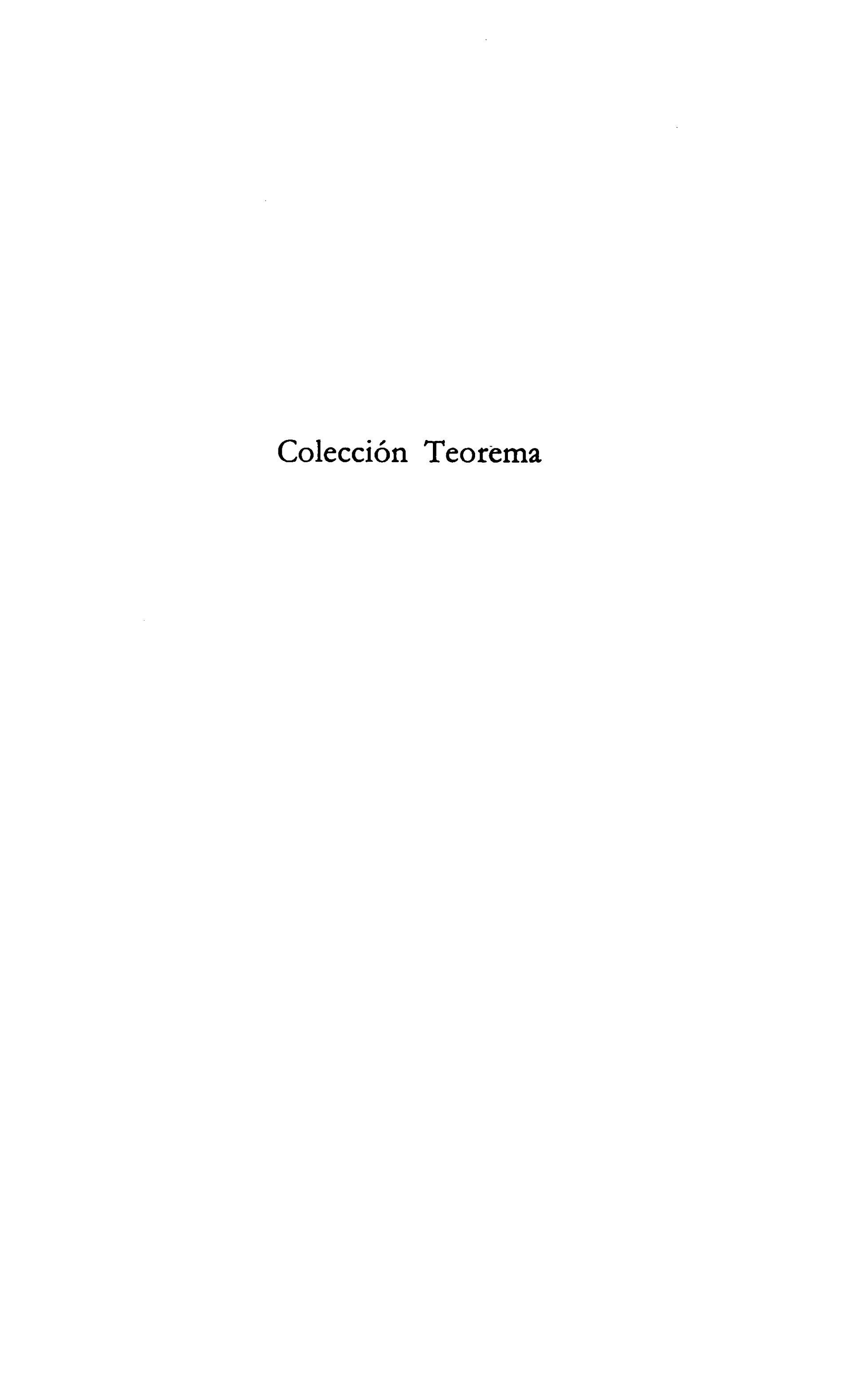
«Todas las continuidades poseen la paradoja de ser absolutamente arbitrarias en sus orígenes y absolutamente ineludibles en su teología. Sabemos esto tan gráficamente gracias a lo que todos nosotros llamamos curiosamente nuestras vidas amorosas que sus equivalentes literarios necesitan poca demostración.» Foucault dice que su forma de considerar la historia de las ideas «permite la introducción, en las mismas raíces del pensamiento, de las ideas de *casualidad, discontinuidad y materialidad*» («The Discourse on Language», incluido en *Archaeology of Knowledge* [Nueva York, 1972], pág. 231.) Lo que más cuesta de todo es ver la contingencia desnuda en la historia de la *filosofía*, aunque sólo fuera porque desde Hegel la historiografía de la filosofía ha sido «progresiva», o (como en la inversión heideggeriana de la explicación del progreso de Hegel) «retrogresiva», pero nunca ha carecido de un sentido de inevitabilidad. Si pudiéramos llegar a ver el deseo de un vocabulario permanente, neutro, ahistórico, conmensurador como si fuera a su vez un fenómeno histórico, entonces quizá podríamos escribir la historia de la filosofía menos dialéctica y sentimentalmente de lo que ha sido posible hasta ahora.

donar la idea de que existe una cosa llamada «método filosófico» o «técnica filosófica» o «el punto de vista filosófico» que permite al filósofo profesional, *ex officio,* tener opiniones interesantes sobre la respetabilidad del psicoanálisis, la legitimidad de ciertas leyes dudosas, la resolución de dilemas morales, la «seriedad» de las escuelas de historiografía o crítica literaria, y otras cuestiones semejantes. Los filósofos tienen muchas veces puntos de vista interesantes sobre estas cuestiones, y su adiestramiento profesional como filósofos suele ser condición necesaria para que tengan los puntos de vista que tienen. Pero esto no significa que los filósofos tengan una clase especial de conocimiento sobre el conocimiento (ni ninguna otra cosa) a partir del cual puedan extraer corolarios relevantes. Las útiles observaciones que pueden hacer sobre los diversos temas que acabamos de mencionar es posible por su familiaridad con los antecedentes históricos de las discusiones sobre temas semejantes, y, lo que es de suma importancia, por el hecho de que las discusiones sobre tales temas van acompañadas de clichés filosóficos desfasados que los demás participantes han encontrado alguna vez en sus lecturas, pero sobre los que los filósofos profesionales saben de memoria los pros y los contras.

La imagen neo-kantiana de la filosofía como una profesión va unidad a la imagen de la «mente» o el «lenguaje» en cuanto espejos de la naturaleza. Por eso puede dar la impresión de que el conductismo epistemológico y el consiguiente rechazo de las imágenes del espejo llevan consigo la afirmación de que no puede o no debe existir tal profesión. Pero esta conclusión no es lógica. Las profesiones pueden sobrevivir a los paradigmas que las hicieron nacer. En cualquier caso, la necesidad de profesores que hayan leído a los grandes filósofos desaparecidos es más que suficiente para garantizar que siga habiendo departamentos de filosofía mientras siga habiendo universidades. El verdadero resultado de la difusión de la pérdida de confianza en las imágenes del espejo sería meramente una «encapsulación» de los problemas creados por estas imágenes dentro de un periodo histórico. No sé si estamos realmente al final de una época. Esto dependerá, supongo, de si Dewey, Wittgenstein y Heidegger llegan o no a ser tomados en serio. Puede ocurrir que las imágenes del espejo y la filosofía sistemática vuelvan a revitalizarse por la acción de algún revolucionario de talento. O puede ocurrir que la imagen del filósofo que ofreció Kant esté a punto de correr la misma suerte que la imagen medieval del sacerdote. Si ocurriera esto, ni los mismos filósofos se tomarán ya en serio la idea de que la filosofía ofrece «fundamentos» o «justificaciones»

para el resto de la cultura, o de que sentencia en las *quaestiones juris* sobre los dominios propios de otras disciplinas.

Pero, ocurra lo que ocurra, no hay peligro de que la filosofía «llegue a su fin». La religión no llegó a su fin con la Ilustración, ni la pintura con el Impresionismo. Aunque el periodo que va de Platón a Nietzsche quede aislado y «distanciado» tal como sugiere Heidegger, y aun cuando la filosofía del siglo XX llegue a parecer una etapa transitoria de apoyo y relleno (como nos lo parece ahora la filosofía del siglo XVI), habrá algo llamado «filosofía» al otro lado de la transición. Efectivamente, aunque los problemas sobre la representación les parezcan a nuestros descendientes tan desfasados como nos parecen a nosotros los problemas del hilemorfismo, la gente seguirá leyendo a Platón, Aristóteles, Descartes, Kant, Hegel, Wittgenstein y Heidegger. No sabemos qué papel ocuparán estos hombres en la conversación de nuestros descendientes. Tampoco sabemos si se mantendrá la distinción entre filosofía sistemática y filosofía edificante. Quizá la filosofía llegue a ser sólo edificante, con lo que la auto-identificación de cada uno en cuanto filósofo se realizará únicamente en términos de los libros que lea y discuta, y no en términos de los problemas que desee resolver. Quizá se encuentre una nueva forma de filosofía sistemática que no tenga absolutamente nada que ver con la epistemología, pero que, sin embargo, haga posible la investigación filosófica normal. Estas especulaciones carecen de fundamento, y nada de lo que he dicho anteriormente hace que una resulte más plausible que otra. Lo único en que yo quisiera insistir es en que el interés moral del filósofo ha de ser que se mantenga la conversación del Occidente, más que el exigir un lugar, dentro de esa conversación, para los problemas tradicionales de la filosofía moderna.



Colección Teorema

Serie mayor

TÍTULOS PUBLICADOS

ABERCROMBIE, N., y otros, *Diccionario de Sociología.*
ACERO, J. J.; BUSTOS, E.; QUESADA, D., *Introducción a la filosofía del lenguaje.* (3.ª ed.).
ANGUERA, María Teresa, *Metodología de la observación en las ciencias humanas* (4.ª ed.).
ATTALI, Jacques, *Los tres mundos.* Para una teoría de la post-crisis.
AYALA, Francisco, *Introducción a las Ciencias Sociales.*
BAUDRILLARD, Jean, *De la seducción* (5.ª ed.).
COHEN, David, *Los psicólogos hablan de Psicología.*
COMREY, Andrew L., *Manual de análisis factorial.*
CHOMSKY, Noam, *Ensayos sobre forma e interpretación.*
DEBRAY, Regis, *Crítica de la razón política.*
DERRIDA, Jacques, *Márgenes de la filosofía.*
EVANS-PRITCHARD, Edward, *Historia del pensamiento antropológico.*
EYSENCK, H. J., *Sexo y personalidad.*
GHISELIN, Michel T., *El triunfo de Darwin.*
HAACK, Susan, *La filosofía de las lógicas.*
HABERMAS, Jürgen, *Teoría de la acción comunicativa: complementos y estudios previos.*
KUTSCHERA, Franz von, *Fundamentos de Ética.*
LYOTARD, Jean-François, *La condición postmoderna* (4.ª ed.).
MORIN, Edgar, *El método. La naturaleza de la naturaleza* (2.ª ed.).
MORIN, Edgar, *El método. La vida de la vida.*
MORIN, Edgar, *El método. El conocimiento del conocimiento.*
RAPHAEL, Bertram, *El computador pensante.* Introducción a la informática para psicólogos y humanistas (2.ª ed.).
RORTY, Richard, *La filosofía y el espejo de la naturaleza* (2.ª ed.).
RUSE, Michael, *La homosexualidad.*
SCHARFSTEIN, Ben-Ami, *Los filósofos y sus vidas.* Para una historia psicológica de la Filosofía.
SEARLE, John, *Actos de habla* (2.ª ed.).
THINÈS, Georges, y LEMPEREUR, Agnés, *Diccionario general de ciencias humanas.*
URMSON, J. O., *Enciclopedia concisa de Filosofía y filósofos* (2.ª ed.).

NOTA FINAL



Le recordamos que este libro ha sido prestado gratuitamente para uso exclusivamente educacional bajo condición de ser destruido una vez leído. Si es así, destrúyalo en forma inmediata.

Súmese como voluntario o donante y promueva este proyecto en su comunidad para que otras personas que no tienen acceso a bibliotecas se vean beneficiadas al igual que usted.

"Es detestable esa avaricia que tienen los que, sabiendo algo, no procuran la transmisión de esos conocimientos".

—*Miguel de Unamuno*

Para otras publicaciones visite:
www.lecturasinegoismo.com
Facebook: Lectura sin Egoísmo
Twitter: @LectSinEgo
o en su defecto escríbanos a:
lecturasinegoismo@gmail.com
Referencia: 444



El presente libro constituye una sensacional «deconstrucción» o desmontaje, desde sus propios supuestos, de la moderna filosofía analítica, como también de la concepción tradicionalmente aceptada de la filosofía.

La idea de que la mente humana es como un espejo que refleja la realidad ha inspirado al pensamiento filosófico desde los griegos. Descartes, Kant y los actuales filósofos analíticos han hecho consistir la tarea del filósofo en limpiar y pulir el espejo de la mente o del lenguaje, para poder establecer así el marco de referencia de todo conocimiento.

Rorty sostiene, sin embargo, que los tres más grandes y más revolucionarios pensadores de nuestro siglo, Wittgenstein, Heidegger y Dewey, han sabido criticar —desde sus respectivos puntos de vista, epistemológico, histórico y social— la validez de la metáfora del espejo. El desarrollo de estas críticas revela que la filosofía analítica se halla en un callejón sin salida. Desde ahora, la filosofía deberá renunciar a su aspiración a presidir el infalible tribunal de la razón pura y contentarse, como ha sugerido Habermas comentando este libro, con el más pragmático y modesto oficio de guardapuestos del saber.

Este libro de Rorty es el único que presenta, por vez primera en la bibliografía actual, un panorama de conjunto y una crítica seria de los grandes pensadores analíticos vivos, como Quine, Davidson, Kuhn o Kripke, en contraste con las corrientes más interesantes de la filosofía continental europea del momento, como la hermenéutica de Gadamer o la dialéctica de Habermas.

«Mucho tiempo habrá de transcurrir —ha escrito Alas Dair Mac Intyre— antes de que vuelva a aparecer una obra como ésta.»





ISBN 84-376-0432-X
9 788437 604329

Colección TEOREMA
serie mayor